



10031519 060602

Atty. Docket: 16335.002

**GENE PRODUCTS DIFFERENTIALLY EXPRESSED IN CANCEROUS COLON CELLS  
AND CORRELATION OF EXPRESSION PATTERNS**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

- [0001]** This application claims the benefit of prior U.S. Provisional Application Serial No. 60/270,959, filed February 21, 2001, which application is incorporated herein by reference.

**FIELD OF THE INVENTION**

- [0002]** This invention relates to methods useful for disease diagnosis by detecting the presence of differences in protein expression levels in patient samples.

**BACKGROUND OF THE INVENTION**

- [0003]** Cancer, like many diseases, is not the result of a single, well-defined cause, but rather can be viewed as several diseases, each caused by different aberrations in informational pathways, that ultimately result in apparently similar pathologic phenotypes. The process of dividing a single general condition into different diseases on the basis of the underlying molecular defects is known as disease stratification. An understanding of the stratification of disease is important in developing therapeutic or preventive drugs, because each layer of a disease may require a different therapy. Implicit is the idea that identification of effective therapeutic or preventive drugs requires not only the identification of the predisposing genes, but also requires understanding the informational pathways in which they operate.
- [0004]** Identification of polynucleotides that correspond to genes that are differentially expressed in cancerous, pre-cancerous, or low metastatic potential cells relative to normal cells of the same tissue type, provides the basis for diagnostic tools, facilitates drug discovery by providing for targets for candidate agents, and further serves to identify therapeutic targets for cancer therapies that are more tailored for the type of cancer to be treated. Identification of differentially expressed gene products also furthers the understanding of the progression and nature of complex diseases such as cancer, and is

essential to identifying the genetic factors that are responsible for the phenotypes associated with development of, for example, the metastatic phenotype. Identification of gene products that are differentially expressed at various stages, and in various types of cancers, can both provide for early diagnostic tests, and further serve as therapeutic targets.

[0005] Early disease diagnosis, especially in diseases such as cancer, is of central importance to halting disease progression, and reducing morbidity. Analysis of a patient's tumor to identify the gene products that are differentially expressed, and administration of therapeutic agent(s) designed to modulate the activity of those differentially expressed gene products, provides the basis for more specific, rationale cancer therapy, which therapy may result in diminished adverse side effects relative to conventional therapies. Furthermore, confirmation that a tumor poses less risk to the patient (*e.g.*, that the tumor is benign) can avoid unnecessary therapies. In short, identification of genes and the encoded gene products that are differentially expressed in cancerous cells can provide the basis of therapeutics, diagnostics, prognostics, therametrics, and the like.

[0006] Clinical trials are critical to the identification and commercial implementation of cancer therapeutics. Clinical trials are very data intensive, time-consuming and costly. A typical cycle for a clinical trial may take several years. For example, designing the trial may take six months, performing the trial may take a year, and analyzing the results may take yet another six months. After years of testing, the results may still be unexpected or difficult to interpret.

[0007] Given the investments in time and money involved, the design of a clinical trial is critical to determining, for example, whether a promising candidate drug -- in which a company has likely already invested a substantial amount of research dollars -- has efficacy in patients. Clinical trials typically are designed to isolate on a single variable, and use a placebo control group as a baseline from which the variable is measured. Observations from a clinical trial attempt to draw conclusions from apparent differences between the control group and the experimental group. These observations, however, do not take into account the multi-variable dynamic nature of the patients individually, or as a

group. Such variations usually increase the variability in the data and require large test populations to deal with the variability in an appropriate statistical manner.

[0008] Where the clinical trial is designed to test the efficacy of a candidate cancer therapy, the selection of patients to provide a homogenous population with respect to the cancer to be treated meets with additional challenges. Conventional diagnosis is largely based upon histopathology and, where available, the use of the few available markers. However, conventional technologies generally do not distinguish tumors at the molecular level. Therefore, the patients selected for inclusion in a clinical trial may have the same tumor type according to conventional technologies, but still represent a heterogenous population due to differences in the tumors at the molecular level that are simply not detected by conventional techniques. This can further confound the correlation of disease phenotypes and treatment efficacy, and the development of diagnostic assays that permit reliable testing with significant predictive value. In addition, useful therapies may be overlooked because the positive results in a subset of patients may be masked by less than positive results of other patients in a heterogenous patient population.

[0009] Thus, there is a need in the field for methods for diagnosis of cancer, as well as methods for classification of cancers at a molecular level to facilitate the selection of subject for clinical trials and the selection of appropriate treatment strategies.

#### SUMMARY OF THE INVENTION

[0010] The present invention relates to methods and compositions useful in diagnosis of colon cancer, design of rational therapy, and the selection of patient populations for the purposes of clinical trials. The invention is based on the discovery that colon tumors of a patient can be classified according to an expression profile of one or more selected genes, which genes are differentially expressed in tumor cells relative to normal cells of the same tissue. Polynucleotides that correspond to the selected differentially expressed genes can be used in diagnostic assays to provide for diagnosis of cancer at the molecular level, and to provide for the basis for rational therapy (*e.g.*, therapy is selected according to the expression pattern of a selected set of genes in the tumor). The gene products encoded by

differentially expressed genes can also serve as therapeutic targets, and candidate agents effective against such targets screened by, for example, analyzing the ability of candidate agents to modulate activity of differentially expressed gene products.

**[0011]** In one aspect, the selected gene(s) for tumor cell (and thus patient) analysis of expression of a gene product encoded by at least one gene selected from at least one of the following groups: 1) Group I, which comprises the genes IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1; and 2) Group II, which comprises the genes IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1.

**[0012]** In another aspect, the invention provides a method for classifying a tumor that shares selected characteristics with respect to a tumor expression profile. In one embodiment, the invention provides a method for classifying a tumor according to an expression profile of one or more genes comprising detecting expression of at least a first Group I gene in a test colon cell sample. Detection of increased expression of the first gene in the test colon cell sample relative to expression of the gene in a control non-cancer cell sample indicates that the tumor is a Group I-type tumor.

**[0013]** In one embodiment, the first Group I gene is an IGF2 gene. In other specific embodiments, the method further comprises detecting expression of a second Group I gene in the test colon cell sample. Detection of increased expression of the first and second genes in the test colon cell sample relative to expression of the first and second genes, respectively, in a control non-cancer cell sample indicates that the tumor is a Group I-type tumor.

**[0014]** In another embodiment, the method further comprises detecting expression of a second and third Group I gene in the test colon cell sample. Detection of increased expression of the first, second, and third genes in the test colon cell sample relative to expression of the first, second, and third genes, respectively, in a control non-cancer cell sample indicates that the tumor is a Group I-type tumor. In other embodiments, the expression of the gene(s) is increased about 1.5-fold, about 2-fold, about 5-fold, or about 10-fold in the test sample relative to the control sample.

- [0015] In another embodiment, the invention provides a method for classifying a tumor according to an expression profile of one or more genes comprising detecting expression of at least a first Group II gene in a test colon cell sample. Detection of increased expression of the first gene in the test colon cell sample relative to expression of the gene in a control non-cancer cell sample indicates that the tumor is a Group II-type tumor.
- [0016] In another embodiment, the first Group II gene is a member of the IFITM family of genes. In other specific embodiments, the method further comprises detecting expression of a second Group II gene in the test colon cell sample. Detection of increased expression of the first and second genes in the test colon cell sample relative to expression of the first and second genes, respectively, in a control non-cancer cell sample indicates that the tumor is a Group II-type tumor. In other embodiments, the expression of the gene(s) is increased about 1.5-fold, about 2-fold, about 5-fold, or about 10-fold in the test sample relative to the control sample. In yet other specific embodiments, the first Group II gene is 1-8U, 1-8D, or 9-27.
- [0017] In another embodiment, the invention provides a method for classifying a tumor according to an expression profile of two or more genes, the method comprising analyzing a test colon cell sample for expression of at least one Group I gene and at least one Group II gene. Detection of increased expression of the at least one Group I gene and the at least one Group II gene in the test cell sample relative to expression of the at least one Group I gene and the at least one Group II gene, respectively, in a control non-cancer cell sample indicates the tumor is a Group I+II-type tumor. In other embodiments, the Group I gene is an IGF2 gene and the Group II gene is a member of the IFITM family of genes. In yet other embodiments, the expression of the genes is increased about 1.5-fold, about 2-fold, about 5-fold, or about 10-fold in the test sample relative to the control sample.
- [0018] In another aspect, the invention provides methods for selection of a patient population having a tumor that shares selected characteristics with respect to a tumor expression profile. This method, referred to herein as "patient stratification," can be used to improve the design of a clinical trial by providing a patient population that is more homogenous with respect to the tumor type that is to be tested for responsiveness to a new

therapy; and in selecting the best therapeutic regiment for a patient in view of an expression profile of the subject's tumor (*e.g.*, rational therapy).

**[0019]** In another aspect, the invention provides a method for selecting an individual for inclusion in a clinical trial, the method comprising the steps of: detecting a level of expression of a gene product in a test colon cell sample or serum obtained from a subject, the gene product being encoded by at least one gene selected from the group consisting of IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1; and comparing the level of expression of the gene product in the test sample to a level of expression in a normal colon cell; wherein detection of a level of expression of the gene product that is significantly higher in the test sample than in a normal cell is a positive indicator for inclusion of the subject in the test population for the clinical trial.

**[0020]** In another aspect the invention provides a method for selecting an individual for inclusion in a clinical trial, the method comprising the steps of: detecting a level of expression of a gene product in a test colon cell sample obtained from a subject, the gene product being encoded by at least one gene selected from the group consisting of: IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1; and comparing the level of expression of the gene product in the test sample to a level of expression in a normal colon cell; wherein detection of a level of expression of the gene product that is significantly higher in the test sample than in a normal cell is a positive indicator for inclusion of the subject in the test population for the clinical trial.

**[0021]** In related aspects the invention provides methods of reducing growth of cancerous colon cells by modulation of expression of one or more gene products corresponding to a gene selected from: 1) Group I, which comprises the genes IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1; and 2) Group II, which comprises the genes IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1. These methods are useful for treating colon cancer.

**[0022]** In another aspect, the present invention provides methods for disease detection by analysis of gene expression. In general, diagnostic and prognostic methods of the invention can involve obtaining a test cell from a subject, e.g., colon cells; detecting the level of expression of any one gene or a selected set of genes in the test cell, where the gene(s) are differentially expressed in a colon tumor cell relative to a normal colon cell; and comparing the expression levels of the gene(s) in the test cell to a control level (e.g., a level of expression in a normal (non-cancerous) colon cell). Detection of a level of expression in the test cell that differs from that found in a normal cell indicates that the test cell is a cancerous cell. The method of the invention permits, for example, detection of a small increase or decrease in gene product production from a gene whose overexpression or underexpression (compared to a reference gene) is associated with cancer or the predisposition for a cancer.

**[0023]** In another aspect the invention provides a method for detecting a cancerous colon cell comprising contacting a sample obtained from a test colon cell with a probe for detection of a gene product of a gene differentially expressed in colon cancer, wherein the gene corresponds to a polynucleotide having a sequence selected from the group consisting of SEQ ID NOS: 1-20, and where contacting is for a time sufficient for binding of the probe to the gene product; and comparing a level of binding of the probe to the sample with a level of probe binding to a control sample obtained from a control colon cell, wherein the control colon cell is of known cancerous state. An increased level of binding of the probe in the test colon cell sample relative to the level of binding in a control sample is indicative of the cancerous state of the test colon cell. In specific embodiments, the probe is a polynucleotide probe and the gene product is nucleic acid. In other specific embodiments, the gene product is a polypeptide. In further embodiments, the gene product or the probe is immobilized on an array.

**[0024]** In another aspect, the invention provides a method for assessing the cancerous phenotype (e.g., metastasis, aberrant cellular proliferation, and the like) of a colon cell comprising detecting expression of a gene product in a test colon cell sample, wherein the gene comprises a sequence selected from the group consisting of SEQ ID NOS: 1-20; and

comparing a level of expression of the gene product in the test colon cell sample with a level of expression of the gene in a control cell sample. Comparison of the level of expression of the gene in the test cell sample relative to the level of expression in the control cell sample is indicative of the cancerous phenotype of the test cell sample. In specific embodiments, detection of expression of the gene is by detecting a level of an RNA transcript in the test cell sample. In other specific embodiments detection of expression of the gene is by detecting a level of a polypeptide in a test sample.

**[0025]** In another aspect, the invention provides a method for suppressing or inhibiting a cancerous phenotype of a cancerous cell, the method comprising introducing into a mammalian cell an antisense polynucleotide for inhibition of expression of a gene comprising a sequence selected from the group consisting of SEQ ID NOS: 1-20. Inhibition of expression of the gene inhibits development of a cancerous phenotype in the cell. In specific embodiments, the cancerous phenotype is metastasis, aberrant cellular proliferation relative to a normal cell, or loss of contact inhibition of cell growth.

**[0026]** In another aspect, the invention provides a method for assessing the tumor burden of a subject, the method comprising detecting a level of a differentially expressed gene product in a test sample from a subject suspected of or having a tumor, the differentially expressed gene product comprising a sequence selected from the group consisting of SEQ ID NOS: 1-20. Detection of the level of the gene product in the test sample is indicative of the tumor burden in the subject.

**[0027]** In another aspect, the invention provides a method for identifying a gene product as a target for a cancer therapeutic, the method comprising contacting a cancerous cell expressing a candidate gene product with an anti-cancer agent, wherein the candidate gene product corresponds to a sequence selected from the group consisting of SEQ ID NOS: 1-20; and analyzing the effect of the anti-cancer agent upon a biological activity of the candidate gene product and upon a cancerous phenotype of the cancerous cell. Modulation of the biological activity of the candidate gene product and modulation of the cancerous phenotype of the cancerous cell indicates the candidate gene product is a target for a cancer therapeutic. In specific embodiments, the cancerous cell is a cancerous colon cell.



In other specific embodiments, the inhibitor is an antisense oligonucleotide. In further embodiments, the cancerous phenotype is aberrant cellular proliferation relative to a normal cell, or colony formation due to loss of contact inhibition of cell growth.

[0028] In another aspect, the invention provides a method for identifying agents that decrease biological activity of a gene product differentially expressed in a cancerous cell, the method comprising contacting a candidate agent with a differentially expressed gene product, the differentially expressed gene product corresponding to a sequence selected from the group consisting of SEQ ID NOS: 1-20; and detecting a decrease in a biological activity of the gene product relative to a level of biological activity of the gene product in the absence of the candidate agent. In specific embodiments, the detecting is by detection of a decrease in expression of the differentially expressed gene product. In other specific embodiments, the gene product is mRNA or cDNA prepared from the mRNA gene product. In further embodiments, the gene product is a polypeptide.

[0029] In all embodiments of the invention, analysis of expression of a gene product of a selected gene can be accomplished by analysis of gene transcription (*e.g.*, by generating cDNA clones from mRNAs isolated from a cell suspected of being cancerous and comparing the number of cDNA clones corresponding to the gene in the sample relative to a number of clones present in a non-cancer cell of the same tissue type), detection of an encoded gene product (*e.g.*, assessing a level of polypeptide encoded by a selected gene present in the test cell suspected of being cancerous relative to a level of the polypeptide in a non-cancer cell of the same tissue type), detection of a biological activity of a gene product encoded by a selected gene, and the like.

[0030] In all embodiments of the invention, comparison of gene product expression of a selected gene in a tumor cell can involve, for example, comparison to an "internal" control cell (*e.g.*, a non-cancer cell of the same tissue type obtained from the same patient from whom the sample suspected of having a tumor cell was obtained), comparison to a control cell analyzed in parallel in the assay (*e.g.*, a non-cancer cell, normally of the same tissue type as the test cell or a cancerous cell, normally of the same tissue type as the test cell), or comparison to a level of gene product expression known to be associated with a normal

cell or a cancerous cell, normally of the same tissue type (e.g., a level of gene product expression is compared to a known level or range of levels of gene product expression for a normal cell or a cancerous cell, which can be provided in the form of, for example, a standard).

[0031] These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

#### BRIEF DESCRIPTION OF THE FIGURES

[0032] Figs. 1-16 are graphs showing the expression profiles of the genes of Group 1.

[0033] Figs. 17-20 are graphs showing the expression profiles of the genes of Group 2.

[0034] In addition to the figures described above, the application also includes Tables 1-3, as well as a Sequence Listing.

#### DETAILED DESCRIPTION OF THE INVENTION

[0035] Before the present invention is described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

[0036] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications and patent applications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0037] It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes a plurality of such

polynucleotides and reference to "the colon cancer cell" includes reference to one or more cells and equivalents thereof known to those skilled in the art, and so forth.

[0038] The publications and applications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

#### DEFINITIONS

[0039] The terms "polynucleotide" and "nucleic acid", used interchangeably herein, refer to a polymeric forms of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, these terms include, but are not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. These terms further include, but are not limited to, mRNA or cDNA that comprise intronic sequences (see, *e.g.*, Niwa *et al.* (1999) *Cell* 99(7):691-702). The backbone of the polynucleotide can comprise sugars and phosphate groups (as may typically be found in RNA or DNA), or modified or substituted sugar or phosphate groups. Alternatively, the backbone of the polynucleotide can comprise a polymer of synthetic subunits such as phosphoramidites and thus can be an oligodeoxynucleoside phosphoramidate or a mixed phosphoramidate-phosphodiester oligomer. Peyrottes *et al.* (1996) *Nucl. Acids Res.* 24:1841-1848; Chaturvedi *et al.* (1996) *Nucl. Acids Res.* 24:2318-2323. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs, uracyl, other sugars, and linking groups such as fluororibose and thioate, and nucleotide branches. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications included in this definition are caps, substitution of one or more of the naturally occurring nucleotides with an analog, and introduction of means for

attaching the polynucleotide to proteins, metal ions, labeling components, other polynucleotides, or a solid support.

[0040] The terms "polypeptide" and "protein", used interchangeably herein, refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. The term includes fusion proteins, including, but not limited to, fusion proteins with a heterologous amino acid sequence, fusions with heterologous and homologous leader sequences, with or without N-terminal methionine residues; immunologically tagged proteins; and the like.

[0041] "Heterologous" means that the materials are derived from different sources (*e.g.*, from different genes, different species, etc.).

[0042] As used herein, the terms "a gene that is differentially expressed in a colon cancer cell," and "a polynucleotide that is differentially expressed in a colon cancer cell" are used interchangeably herein, and generally refer to a polynucleotide that represents or corresponds to a gene that is differentially expressed in a cancerous colon cell when compared with a cell of the same cell type that is not cancerous, *e.g.*, mRNA is found at levels at least about 25%, at least about 50% to about 75%, at least about 90%, at least about 1.5-fold, at least about 2-fold, at least about 5-fold, at least about 10-fold, or at least about 50-fold or more, different (*e.g.*, higher or lower). The comparison can be made in tissue, for example, if one is using in situ hybridization or another assay method that allows some degree of discrimination among cell types in the tissue. The comparison may also or alternatively be made between cells removed from their tissue source.

[0043] "Differentially expressed polynucleotide" as used herein refers to a nucleic acid molecule (RNA or DNA) comprising a sequence that represents a differentially expressed gene, *e.g.*, the differentially expressed polynucleotide comprises a sequence (*e.g.*, an open reading frame encoding a gene product; a non-coding sequence) that uniquely identifies a differentially expressed gene so that detection of the differentially expressed polynucleotide in a sample is correlated with the presence of a differentially expressed gene in a sample. "Differentially expressed polynucleotides" is also meant to encompass

fragments of the disclosed polynucleotides, *e.g.*, fragments retaining biological activity, as well as nucleic acids homologous, substantially similar, or substantially identical (*e.g.*, having about 90% sequence identity) to the disclosed polynucleotides.

[0044] "Corresponds to" or "represents" when used in the context of, for example, a polynucleotide or sequence that "corresponds to" or "represents" a gene means that a sequence of the polynucleotide is present in the gene or in a cDNA produced from an mRNA that is a product of the gene. The polynucleotide may be wholly present within an exon of a genomic sequence of the gene, or different portions of the sequence of the polynucleotide may be present in different exons (*e.g.*, such that the contiguous polynucleotide sequence is present in an mRNA, either pre- or post-splicing, that is an expression product of the gene). In some embodiments, the polynucleotide may represent or correspond to a gene that is modified in a cancerous cell relative to a normal cell. For example, the gene in the cancerous cell may be modified by insertion of an endogenous retrovirus, a transposable element, or other naturally occurring or non-naturally occurring nucleic acid. In such cases, the polynucleotide may include sequences of both the native gene (*e.g.*, the gene without the heterologous sequence) and the inserted, non-native sequence.

[0045] "Diagnosis" as used herein generally includes determination of a subject's susceptibility to a disease or disorder, determination as to whether a subject is presently affected by a disease or disorder, prognosis of a subject affected by a disease or disorder (*e.g.*, identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy), and use of therametrics (*e.g.*, monitoring a subject's condition to provide information as to the effect or efficacy of therapy).

[0046] As used herein, the term "a polypeptide associated with colon cancer" refers to a polypeptide encoded by a polynucleotide that is differentially expressed in a colon cancer cell.

[0047] The term "biological sample" encompasses a variety of sample types obtained from an organism and can be used in a diagnostic or monitoring assay. The term encompasses blood and other liquid samples of biological origin, solid tissue samples, such as a biopsy

specimen or tissue cultures or cells derived therefrom and the progeny thereof. The term encompasses samples that have been manipulated in any way after their procurement, such as by treatment with reagents, solubilization, or enrichment for certain components. The term encompasses a clinical sample, and also includes cells in cell culture, cell supernatants, cell lysates, serum, plasma, biological fluids, and tissue samples.

**[0048]** The terms "treatment", "treating", "treat" and the like are used herein to generally refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete stabilization or cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease in a mammal, particularly a human, and includes: (a) preventing the disease or symptom from occurring in a subject which may be predisposed to the disease or symptom but has not yet been diagnosed as having it; (b) inhibiting the disease symptom, i.e., arresting its development; or (c) relieving the disease symptom, i.e., causing regression of the disease or symptom.

**[0049]** The terms "individual," "subject," "host," and "patient," used interchangeably herein and refer to any mammalian subject for whom diagnosis, treatment, or therapy is desired, particularly humans. Other subjects may include cattle, dogs, cats, guinea pigs, rabbits, rats, mice, horses, and the like.

**[0050]** As used herein the term "isolated" refers to a polynucleotide, a polypeptide, an antibody, or a host cell that is in an environment different from that in which the polynucleotide, the polypeptide, the antibody, or the host cell naturally occurs. A polynucleotide, a polypeptide, an antibody, or a host cell which is isolated is generally substantially purified. As used herein, the term "substantially purified" refers to a compound (*e.g.*, either a polynucleotide or a polypeptide or an antibody) that is removed from its natural environment and is at least 60% free, preferably 75% free, and most preferably 90% free from other components with which it is naturally associated. Thus, for example, a composition containing A is "substantially free of" B when at least 85% by weight of the total A+B in the composition is A. Preferably, A comprises at least about

90% by weight of the total of A+B in the composition, more preferably at least about 95% or even 99% by weight.

- [0051] A "host cell", as used herein, refers to a microorganism or a eukaryotic cell or cell line cultured as a unicellular entity which can be, or has been, used as a recipient for a recombinant vector or other transfer polynucleotides, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.
- [0052] The terms "cancer", "neoplasm", "tumor", and "carcinoma", are used interchangeably herein to refer to cells which exhibit relatively autonomous growth, so that they exhibit an aberrant growth phenotype characterized by a significant loss of control of cell proliferation. In general, cells of interest for detection or treatment in the present application include precancerous (*e.g.*, benign), malignant, pre-metastatic, metastatic, and non-metastatic cells. Detection of cancerous cell is of particular interest.
- [0053] The term "normal" as used in the context of "normal cell," is meant to refer to a cell of an untransformed phenotype or exhibiting a morphology of a non-transformed cell of the tissue type being examined.
- [0054] "Cancerous phenotype" generally refers to any of a variety of biological phenomena that are characteristic of a cancerous cell, which phenomena can vary with the type of cancer. The cancerous phenotype is generally identified by abnormalities in, for example, cell growth or proliferation (*e.g.*, uncontrolled growth or proliferation), regulation of the cell cycle, cell mobility, or cell-cell interaction.
- [0055] "Therapeutic target" generally refers to a gene or gene product that, upon modulation of its activity (*e.g.*, by modulation of expression, biological activity, and the like), can provide for modulation of the cancerous phenotype.
- [0056] As used throughout, "modulation" is meant to refer to an increase or a decrease in the indicated phenomenon (*e.g.*, modulation of a biological activity refers to an increase in a biological activity or a decrease in a biological activity).

[0057] As used herein a "Group I type tumor" is a tumor comprising cells that, relative to a non-cancer cell of the same tissue type, exhibit increased expression of a gene product encoded by at least one or more of the following genes: IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1.

[0058] As used herein a "Group II type tumor" is a tumor comprising cells that, relative to a non-cancer cell of the same tissue type, exhibit increased expression of a gene product encoded by at least one or more of the following genes: IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1.

[0059] As used herein a "Group I+II type tumor" is a tumor comprising cells that, relative to a non-cancer cell of the same tissue type, exhibit increased expression of 1) a gene product encoded by at least one or more of the following genes: IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1; and a gene product encoded by at least one or more of the following genes 2) IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1.

#### Overview

[0060] The invention is based on the discovery of genes differentially expressed in cancerous colon cells relative to normal cells, particularly metastatic or pre-metastatic cancerous colon cells relative to normal cells of the same tissue type. The genes of particular interest are those described in the Examples below. The invention is further based on the discovery that colon tumors can be classified according to the expression pattern of one or more of genes, and that patients can thus be classified and diagnosed, and therapy selected accordingly, according to these expression patterns. The gene(s) for analysis of expression of a gene product encoded by at least one gene selected from at least one of the following groups: 1) Group I, which comprises the genes IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1; and 2) Group II, which comprises the genes IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1. . A tumor can then be classified as



a Group I-type, Group II-type, or Group I+II-type tumor based on the expression profile of the tumor. The expression patterns associated with colon cancer, and which provide the basis for tumor classification and patient stratification, are described in the Examples below.

[0061] The methods of the invention can be carried out using any suitable probe for detection of a gene product that is differentially expressed in colon cancer cells. For example, mRNA (or cDNA generated from mRNA) expressed from a differentially expressed gene can be detected using polynucleotide probes. In another example, the differentially expressed gene product is a polypeptide, which polypeptides can be detected using, for example, antibodies that specifically bind such polypeptides or an antigenic portion thereof.

[0062] The invention will now be described in more detail.

#### Polynucleotide Compositions

[0063] The present invention provides isolated polynucleotides that represent genes that are differentially expressed in colon cancer cells. The polynucleotides, as well as polypeptides encoded thereby, find use in a variety of therapeutic and diagnostic methods.

[0064] The scope of the invention with respect to polynucleotide compositions useful in the methods described herein includes, but is not necessarily limited to, polynucleotides having a sequence set forth in any one of the polynucleotide sequences provided herein; polynucleotides obtained from the biological materials described herein or other biological sources (particularly human sources) by hybridization under stringent conditions (particularly conditions of high stringency); genes corresponding to the provided polynucleotides; cDNAs corresponding to the provided polynucleotides; variants of the provided polynucleotides and their corresponding genes, particularly those variants that retain a biological activity of the encoded gene product (*e.g.*, a biological activity ascribed to a gene product corresponding to the provided polynucleotides as a result of the assignment of the gene product to a protein family(ies) and/or identification of a functional

domain present in the gene product). Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here. "Polynucleotide" and "nucleic acid" as used herein with reference to nucleic acids of the composition is not intended to be limiting as to the length or structure of the nucleic acid unless specifically indicated.

[0065] The invention features polynucleotides that represent genes that are expressed in human tissue, specifically human colon tissue, particularly polynucleotides that are differentially expressed in colon cancer cells. Nucleic acid compositions described herein of particular interest comprise a sequence set forth in any one of the polynucleotide sequences provided herein or an identifying sequence thereof. An "identifying sequence" is a contiguous sequence of residues at least about 10 nt to about 20 nt in length, usually at least about 50 nt to about 100 nt in length, that uniquely identifies a polynucleotide sequence, its complements and degenerate variants thereof, *e.g.*, exhibits less than 90%, usually less than about 80% to about 85% sequence identity to any contiguous nucleotide sequence of more than about 20 nt. Thus, the subject nucleic acid compositions include full-length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of the polynucleotide sequences provided herein.

[0066] The polynucleotides useful in the methods described herein also include polynucleotides having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under high stringency conditions, for example, at 50°C or higher and 0.1XSSC (9 mM saline/0.9 mM sodium citrate). Hybridization methods and conditions are well known in the art, see, *e.g.*, USPN 5,707,829. Nucleic acids that are substantially identical to the provided polynucleotide sequences, *e.g.* allelic variants, genetically altered versions of the gene, *etc.*, bind to the provided polynucleotide sequences under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any

species, *e.g.* primate species, particularly human; rodents, such as rats and mice; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.*

[0067] In one embodiment, hybridization is performed using at least 15 contiguous nucleotides (nt) of at least one of the polynucleotide sequences provided herein. That is, when at least 15 contiguous nt of one of the disclosed polynucleotide sequences is used as a probe, the probe will preferentially hybridize with a nucleic acid comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids that uniquely hybridize to the selected probe. Probes from more than one polynucleotide sequences provided herein can hybridize with the same nucleic acid if the cDNA from which they were derived corresponds to one mRNA. Probes of more than 15 nt can be used, *e.g.*, probes of a size within the range of about 18 nt, 25 nt, 50 nt, 75 nt or 100 nt, but in general about 15 nt represents sufficient sequence for unique identification.

[0068] Polynucleotides contemplated by the invention also include naturally occurring variants of the nucleotide sequences (*e.g.*, degenerate variants (*e.g.*, sequences that encode the same polypeptides but, due to the degenerate nature of the genetic code, different in nucleotide sequence), allelic variants, *etc.*). Variants of the polynucleotides contemplated by the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the polynucleotides described herein can be identified where the allelic variant exhibits at most about 25-30% base pair (bp) mismatches relative to the selected polynucleotide probe. In general, allelic variants contain 15-25% bp mismatches, and can contain as little as even 5-15%, or 2-5%, or 1-2% bp mismatches, as well as a single bp mismatch.

[0069] The invention also encompasses homologs corresponding to any one of the polynucleotide sequences provided herein, where the source of homologous genes can be any mammalian species, *e.g.*, primate species, particularly human; rodents, such as rats; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.* Between mammalian species, *e.g.*, human and mouse, homologs generally have substantial sequence similarity, *e.g.*, at least 75% sequence identity, usually at least 90%, more usually at least 95%

between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about 18 contiguous nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as gapped BLAST, described in Altschul, et al. *Nucleic Acids Res.* (1997) 25:3389-3402, or TeraBLAST available from TimeLogic Corp. (Crystal Bay, Nevada).

[0070] In general, variants of the polynucleotides described herein have a sequence identity greater than at least about 65%, preferably at least about 75%, more preferably at least about 85%, and can be greater than at least about 90%, 95%, 96%, 97%, 98%, 99% or more as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular). For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm. Global DNA sequence identity must be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty, 12; and gap extension penalty, 1.

[0071] The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein (*e.g.*, in diagnosis, as a unique identifier of a differentially expressed gene of interest, *etc.*). The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide. mRNA species can also exist with both exons and introns, where the introns may be removed by alternative splicing. Furthermore it should be noted that different species of mRNAs encoded by the same genomic sequence

can exist at varying levels in a cell, and detection of these various levels of mRNA species can be indicative of differential expression of the encoded gene product in the cell.

**[0072]** A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression.

**[0073]** The nucleic acid compositions of the subject invention can encode all or a part of the naturally-occurring polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* Isolated polynucleotides and polynucleotide fragments contemplated by the invention comprise at least about 10, about 15, about 20, about 35, about 50, about 75, about 100, about 150 to about 200, about 250 to about 300, or about 350 contiguous nt selected from the polynucleotide provided herein. For the most part, fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and up to at least about 50 contiguous nt in length or more. In a preferred embodiment, the polynucleotide molecules comprise a contiguous sequence of at least 12 nt selected from any one of the polynucleotide sequences provided herein.

**[0074]** Probes specific to the polynucleotides described herein can be generated using the polynucleotide sequences disclosed herein. The probes are preferably at least about a 12 nt, 15 nt, 16 nt, 18 nt, 20 nt, 22 nt, 24 nt, or 25 nt fragment of a corresponding contiguous sequence any one of the polynucleotide sequences provided herein, and can be less than 10kb, 5kb, 4kb, 3kb, 2 kb, 1 kb, 0.5 kb, 0.1 kb, or 0.05 kb in length. The probes can be

synthesized chemically or can be generated from longer polynucleotides using restriction enzymes. The probes can be labeled, for example, with a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of any one of the polynucleotide sequences provided herein. More preferably, probes are designed based on a contiguous sequence of one of the subject polynucleotides that remain unmasked following application of a masking program for masking low complexity (*e.g.*, XBLAST, RepeatMasker, etc.) to the sequence., *i.e.*, one would select an unmasked region, as indicated by the polynucleotides outside the poly-n stretches of the masked sequence produced by the masking program.

[0075] The polynucleotides of the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the polynucleotides, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

[0076] The polynucleotides described herein can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the polynucleotides can be regulated by their own or by other regulatory sequences known in the art. The polynucleotides can be introduced into suitable host cells using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

[0077] The nucleic acid compositions described herein can be used to, for example, produce polypeptides, as probes for the detection of mRNA in biological samples (*e.g.*, extracts of human cells) or cDNA produced from such samples, to generate additional copies of the polynucleotides, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes

described herein can be used to, for example, determine the presence or absence of any one of the polynucleotide provided herein or variants thereof in a sample. These and other uses are described in more detail below.

#### Polypeptides and Variants Thereof

- [0078] The present invention further provides polypeptides encoded by polynucleotides that represent genes that are differentially expressed in colon cancer cells. Such polypeptides are referred to herein as "polypeptides associated with colon cancer." The polypeptides can be used to generate antibodies specific for a polypeptide associated with colon cancer, which antibodies are in turn useful in diagnostic methods, prognostics methods, therametric methods, and the like as discussed in more detail herein. Polypeptides are also useful as targets for therapeutic intervention, as discussed in more detail herein.
- [0079] The polypeptides contemplated by the invention include those encoded by the disclosed polynucleotides and the genes to which these polynucleotides correspond, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed polynucleotides. Thus, the invention includes within its scope a polypeptide encoded by a polynucleotide having the sequence of any one of the polynucleotide sequences provided herein, or a variant thereof.
- [0080] In general, the term "polypeptide" as used herein refers to both the full length polypeptide encoded by the recited polynucleotide, the polypeptide encoded by the gene represented by the recited polynucleotide, as well as portions or fragments thereof. "Polypeptides" also includes variants of the naturally occurring proteins, where such variants are homologous or substantially similar to the naturally occurring protein, and can be of an origin of the same or different species as the naturally occurring protein (*e.g.*, human, murine, or some other species that naturally expresses the recited polypeptide, usually a mammalian species). In general, variant polypeptides have a sequence that has at least about 80%, usually at least about 90%, and more usually at least about 98% sequence identity with a differentially expressed polypeptide described herein, as measured by

TeraBLAST using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, *i.e.*, the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

[0081] The invention also encompasses homologs of the disclosed polypeptides (or fragments thereof) where the homologs are isolated from other species, *i.e.* other animal or plant species, where such homologs, usually mammalian species, *e.g.* rodents, such as mice, rats; domestic animals, *e.g.*, horse, cow, dog, cat; and humans. By "homolog" is meant a polypeptide having at least about 35%, usually at least about 40% and more usually at least about 60% amino acid sequence identity to a particular differentially expressed protein as identified above, where sequence identity is determined using the TeraBLAST algorithm, with the parameters described *supra*.

[0082] In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment, *e.g.* are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition that is enriched for the protein as compared to a control. As such, purified polypeptide is provided, where by purified is meant that the protein is present in a composition that is substantially free of non-differentially expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of non-differentially expressed polypeptides.

[0083] Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/hydrophilicity, and/or steric bulk of the amino acid substituted.



[0084] Variants can be designed so as to retain or have enhanced biological activity of a particular region of the protein (e.g., a functional domain and/or, where the polypeptide is a member of a protein family, a region associated with a consensus sequence). Selection of amino acid alterations for production of variants can be based upon the accessibility (interior vs. exterior) of the amino acid (see, e.g., Go *et al.*, *Int. J. Peptide Protein Res.* (1980) 15:211), the thermostability of the variant polypeptide (see, e.g., Querol *et al.*, *Prot. Eng.* (1996) 9:265), desired glycosylation sites (see, e.g., Olsen and Thomsen, *J. Gen. Microbiol.* (1991) 137:579), desired disulfide bridges (see, e.g., Clarke *et al.*, *Biochemistry* (1993) 32:4322; and Wakarchuk *et al.*, *Protein Eng.* (1994) 7:1379), desired metal binding sites (see, e.g., Toma *et al.*, *Biochemistry* (1991) 30:97, and Haezebrouck *et al.*, *Protein Eng.* (1993) 6:643), and desired substitutions with in proline loops (see, e.g., Masul *et al.*, *Appl. Env. Microbiol.* (1994) 60:3579). Cysteine-depleted muteins can be produced as disclosed in USPN 4,959,314.

[0085] Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 aa to at least about 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a polynucleotide having a sequence of any one of the polynucleotide sequences provided herein, or a homolog thereof. The protein variants described herein are encoded by polynucleotides that are within the scope of the invention. The genetic code can be used to select the appropriate codons to construct the corresponding variants.

#### Antibodies and Other Polypeptide or Polynucleotide Binding Molecules

[0086] The present invention further provides antibodies, which may be isolated antibodies, that are specific for a polypeptide encoded by a polynucleotide described herein and/or a polypeptide of a gene that corresponds to a polynucleotide described herein. Antibodies can be provided in a composition comprising the antibody and a buffer

and/or a pharmaceutically acceptable excipient. Antibodies specific for a polypeptide associated with colon cancer are useful in a variety of diagnostic and therapeutic methods, as discussed in detail herein.

[0087] Gene products, including polypeptides, mRNA (particularly mRNAs having distinct secondary and/or tertiary structures), cDNA, or complete gene, can be prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. For polynucleotides to which a corresponding gene has not been assigned, this provides an additional method of identifying the corresponding gene. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native protein in a cell or tissue preparation or in a cell-free extract of an in vitro expression system.

[0088] Methods for production of antibodies that specifically bind a selected antigen are well known in the art. Immunogens for raising antibodies can be prepared by mixing an antigen (*e.g.*, polypeptide) with an adjuvant, and/or by making fusion proteins with larger immunogenic proteins. Antigens (*e.g.*, polypeptides) can also be covalently linked to other larger immunogenic proteins, such as keyhole limpet hemocyanin. Immunogens are typically administered intradermally, subcutaneously, or intramuscularly to experimental animals such as rabbits, sheep, and mice, to generate antibodies. Monoclonal antibodies can be generated by isolating spleen cells and fusing myeloma cells to form hybridomas. Alternatively, a polynucleotide encoding an antigen of interest is administered directly, such as by intramuscular injection, and expressed in vivo. The expressed protein antigen generates a variety of protein-specific immune responses, including production of antibodies, comparable to administration of the protein.

[0089] Preparations of polyclonal and monoclonal antibodies specific for an antigen (*e.g.*, polypeptide) are made using standard methods known in the art. For example, the antibodies can be produced so as to specifically bind to epitopes present in the polypeptides encoded by polynucleotides disclosed in the Sequence Listing. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. Epitopes that

involve non-contiguous amino acids may require a longer polypeptide, e.g., at least 15, 25, or 50 amino acids. Antibodies that specifically bind to human polypeptides encoded by the provided polypeptides should provide a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in Western blots or other immunochemical assays. In one embodiment, antibodies that specifically bind polypeptides contemplated by the invention do not bind to other proteins in immunochemical assays at detectable levels and can immunoprecipitate the specific polypeptide from solution.

[0090] The invention also contemplates naturally occurring antibodies. For example, serum antibodies to a polypeptide of interest in a human population can be purified by methods well known in the art, e.g., by passing antiserum over a column to which the corresponding selected polypeptide or fusion protein is bound. The bound antibodies can then be eluted from the column, for example using a buffer with a high salt concentration.

[0091] The invention also contemplates genetically engineered antibodies (e.g., chimeric antibodies, humanized antibodies, human antibodies produced by a transgenic animal (e.g., a transgenic mouse such as the Xenomous<sup>TM</sup>), antibody derivatives (e.g., single chain antibodies, antibody fragments (e.g., Fab, etc.)), according to methods well known in the art.

[0092] The invention also contemplates other molecules that can specifically bind a polynucleotide or polypeptide of the invention. Examples of such molecules include, but are not necessarily limited to, single-chain binding proteins (e.g., mono- and multi-valent single chain antigen binding proteins (see, e.g., U.S. Patent Nos. 4,704,692; 4,946,778; 4,946,778; 6,027,725; 6,121,424)), oligonucleotide-based synthetic antibodies (e.g., oligobodies (see, e.g., Radrizzani *et al.*, *Medicina* (B Aires) (1999) 59:753-8; Radrizzani *et al.*, *Medicina* (B Aires) (2000) 60(Suppl 2):55-60)), aptamers (see, e.g., Gening *et al.*, *Biotechniques* (2001) 3:828, 830, 832, 834; Cox and Ellington, *Bioorg. Med. Chem.* (2001) 9:2525-31), and the like.

## Computer-Related Embodiments

[0093] In general, a library of polynucleotides is a collection of sequence information, which information is provided in either biochemical form (*e.g.*, as a collection of polynucleotide molecules), or in electronic form (*e.g.*, as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The sequence information of the polynucleotides can be used in a variety of ways, *e.g.*, as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (*e.g.*, cell type markers), and/or as markers of a given disease or disease state. For example, in the instant case, the sequences of polynucleotides and polypeptides corresponding to genes differentially expressed in cancer, particular in colon cancer, as well as the nucleic acid and amino acid sequences of the genes themselves, can be provided in electronic form in a computer database.

**[0094]** In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (*e.g.*, a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either overexpressed or underexpressed in a cancerous colon cell affected by cancer relative to a normal (*i.e.*, substantially disease-free) colon cell.

**[0095]** The nucleotide sequence information of the library can be embodied in any suitable form, *e.g.*, electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file (or, in biochemical form, a collection of nucleic acid molecules) that contains the representative nucleotide sequences of genes that are differentially expressed (*e.g.*, overexpressed or underexpressed) as between, for example, i) a cancerous cell and a normal cell; ii) a cancerous cell and a dysplastic cell; iii) a cancerous cell and a cell affected by a disease or condition other than cancer; iv) a metastatic cancerous cell and a normal cell and/or non-metastatic cancerous cell; v) a malignant cancerous cell and a non-malignant cancerous cell (or a normal cell) and/or vi) a dysplastic cell relative to a normal cell. Other

combinations and comparisons of cells affected by various diseases or stages of disease will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acids that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

[0096] The polynucleotide libraries of the subject invention generally comprise sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence of any of sequence described herein. By plurality is meant at least 2, usually at least 3 and can include up to all of the sequences described herein. The length and number of polynucleotides in the library will vary with the nature of the library, *e.g.*, if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, etc.

[0097] Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, *e.g.* the nucleic acid sequences of any of the polynucleotides of the sequences described herein, can be recorded on computer readable media, *e.g.* any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

[0098] One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the

stored information. A variety of data processor programs and formats can be used for storage, *e.g.* word processing text file, database format, *etc.* In addition to the sequence information, electronic versions of libraries comprising one or more sequence described herein can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (*e.g.*, searchable files, executable files, *etc.*, including, but not limited to, for example, search program software, *etc.*).

[0099] By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the gapped BLAST (Altschul *et al.*, *Nucleic Acids Res.* (1997) 25:3389-3402) and BLAZE (Brutlag *et al.*, *Comp. Chem.* (1993) 17:203) search algorithms on a Sybase system, or the TeraBLAST (TimeLogic, Crystal Bay, Nevada) program optionally running on a specialized computer platform available from TimeLogic, can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

[00100] As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

[00101] "Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, *e.g.* MacPattern (EMBL), TeraBLAST (TimeLogic),

BLASTN and BLASTX (NCBI). A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nt. A variety of comparing means can be used to accomplish comparison of sequence information from a sample (e.g., to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art.

[00102] A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

[00103] A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile. A gene expression profile can be generated from, for example, a cDNA library prepared from mRNA isolated from a test cell suspected of being cancerous or pre-cancerous, comparing the sequences or partial sequences of the clones against the sequences in an electronic database, where the sequences of the electronic database represent genes differentially expressed in a cancerous cell, e.g., a cancerous colon cell. The number of clones having a sequence that has substantial similarity to a sequence that represents a gene differentially expressed in a

cancerous cell is then determined, and the number of clones corresponding to each of such genes is determined. An increased number of clones that correspond to differentially expressed gene is present in the cDNA library of the test cell (relative to, for example, the number of clones expected in a cDNA of a normal cell) indicates that the test cell is cancerous.

**[00104]** As discussed above, the "library" as used herein also encompasses biochemical libraries of the polynucleotides of the sequences described herein, *e.g.*, collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, *e.g.*, a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (*i.e.*, an array) and the like. Of particular interest are nucleic acid arrays in which one or more of the genes described herein is represented by a sequence on the array. By array is meant an article of manufacture that has at least a substrate with at least two distinct nucleic acid targets on one of its surfaces, where the number of distinct nucleic acids can be considerably higher, typically being at least 10 nt, usually at least 20 nt and often at least 25 nt. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

**[00105]** In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the polypeptides of the library will represent at least a portion of the polypeptides encoded by a gene corresponding to a sequence described herein.

#### Diagnostic and Other Methods Involving Detection of Differentially Expressed Genes

**[00106]** The present invention provides methods of using the polynucleotides described herein. In specific non-limiting embodiments, the methods are useful for detecting colon cancer cells, facilitating diagnosis of cancer and the severity of a cancer (*e.g.*, tumor grade, tumor burden, and the like) in a subject, facilitating a determination of the prognosis of a subject, and assessing the responsiveness of the subject to therapy (*e.g.*, by providing a



measure of therapeutic effect through, for example, assessing tumor burden during or following a chemotherapeutic regimen). Detection can be based on detection of a polynucleotide that is differentially expressed in a colon cancer cell, and/or detection of a polypeptide encoded by a polynucleotide that is differentially expressed in a colon cancer cell ("a polypeptide associated with colon cancer"). The detection methods of the invention can be conducted *in vitro* or *in vivo*, on isolated cells, or in whole tissues or a bodily fluid, *e.g.*, blood, plasma, serum, urine, and the like).

[00107] In general, methods of the invention involving detection of a gene product (*e.g.*, mRNA, cDNA generated from such mRNA, and polypeptides) involve contacting a sample with a probe specific for the gene product of interest. "Probe" as used herein in such methods is meant to refer to a molecule that specifically binds a gene product of interest (*e.g.*, the probe binds to the target gene product with a specificity sufficient to distinguish binding to target over non-specific binding to non-target (background) molecules). "Probes" include, but are not necessarily limited to, nucleic acid probes (*e.g.*, DNA, RNA, modified nucleic acid, and the like), antibodies (*e.g.*, antibodies, antibody fragments that retain binding to a target epitope, single chain antibodies, and the like), or other polypeptide, peptide, or molecule (*e.g.*, receptor ligand) that specifically binds a target gene product of interest.

[00108] The probe and sample suspected of having the gene product of interest are contacted under conditions suitable for binding of the probe to the gene product. For example, contacting is generally for a time sufficient to allow binding of the probe to the gene product (*e.g.*, from several minutes to a few hours), and at a temperature and conditions of osmolarity and the like that provide for binding of the probe to the gene product at a level that is sufficiently distinguishable from background binding of the probe (*e.g.*, under conditions that minimize non-specific binding). Suitable conditions for probe-target gene product binding can be readily determined using controls and other techniques available and known to one of ordinary skill in the art.

[00109] In this embodiment, the probe can be an antibody or other polypeptide, peptide, or molecule (*e.g.*, receptor ligand) that specifically binds a target polypeptide of interest.

[00110] The detection methods can be provided as part of a kit. Thus, the invention further provides kits for detecting the presence and/or a level of a polynucleotide that is differentially expressed in a colon cancer cell (*e.g.*, by detection of an mRNA encoded by the differentially expressed gene of interest), and/or a polypeptide encoded thereby, in a biological sample. Procedures using these kits can be performed by clinical laboratories, experimental laboratories, medical practitioners, or private individuals. The kits of the invention for detecting a polypeptide encoded by a polynucleotide that is differentially expressed in a colon cancer cell comprise a moiety that specifically binds the polypeptide, which may be a specific antibody. The kits of the invention for detecting a polynucleotide that is differentially expressed in a colon cancer cell comprise a moiety that specifically hybridizes to such a polynucleotide. The kit may optionally provide additional components that are useful in the procedure, including, but not limited to, buffers, developing reagents, labels, reacting surfaces, means for detection, control samples, standards, instructions, and interpretive information.

Detecting a polypeptide encoded by a polynucleotide that is differentially expressed in a colon cancer cell

[00111] In some embodiments, methods are provided for a colon cancer cell by detecting in the cell a polypeptide encoded by a gene differentially expressed in a colon cancer cell. Any of a variety of known methods can be used for detection, including, but not limited to, immunoassay, using antibody specific for the encoded polypeptide, *e.g.*, by enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and the like; and functional assays for the encoded polypeptide, *e.g.*, binding activity or enzymatic activity.

[00112] For example, an immunofluorescence assay can be easily performed on cells without first isolating the encoded polypeptide. The cells are first fixed onto a solid support, such as a microscope slide or microtiter well. This fixing step can permeabilize the cell membrane. The permeabilization of the cell membrane permits the polypeptide-specific probe (*e.g.*, antibody) to bind. Alternatively, where the polypeptide is secreted or membrane-bound, or is otherwise accessible at the cell-surface (*e.g.*, receptors, and other

molecule stably-associated with the outer cell membrane or otherwise stably associated with the cell membrane, such permeabilization may not be necessary.

[00113] Next, the fixed cells are exposed to an antibody specific for the encoded polypeptide. To increase the sensitivity of the assay, the fixed cells may be further exposed to a second antibody, which is labeled and binds to the first antibody, which is specific for the encoded polypeptide. Typically, the secondary antibody is detectably labeled, e.g., with a fluorescent marker. The cells which express the encoded polypeptide will be fluorescently labeled and easily visualized under the microscope. See, for example, Hashido *et al.* (1992) *Biochem. Biophys. Res. Comm.* 187:1241-1248.

[00114] As will be readily apparent to the ordinarily skilled artisan upon reading the present specification, the detection methods and other methods described herein can be readily varied. Such variations are within the intended scope of the invention. For example, in the above detection scheme, the probe for use in detection can be immobilized on a solid support, and the test sample contacted with the immobilized probe. Binding of the test sample to the probe can then be detected in a variety of ways, e.g., by detecting a detectable label bound to the test sample to facilitate detected of test sample-immobilized probe complexes.

[00115] The present invention further provides methods for detecting the presence of and/or measuring a level of a polypeptide in a biological sample, which polypeptide is encoded by a polynucleotide that represents a gene differentially expressed in cancer, particularly in a colon cancer cell, using a probe specific for the encoded polypeptide. In this embodiment, the probe can be an antibody or other polypeptide, peptide, or molecule (e.g., receptor ligand) that specifically binds a target polypeptide of interest.

[00116] The methods generally comprise: a) contacting the sample with an antibody specific for a differentially expressed polypeptide in a test cell; and b) detecting binding between the antibody and molecules of the sample. The level of antibody binding (either qualitative or quantitative) indicates the cancerous state of the cell. For example, where the differentially expressed gene is increased in cancerous cells, detection of an increased

level of antibody binding to the test sample relative to antibody binding level associated with a normal cell indicates that the test cell is cancerous.

[00117] Suitable controls include a sample known not to contain the encoded polypeptide; and a sample contacted with an antibody not specific for the encoded polypeptide, e.g., an anti-idiotypic antibody. A variety of methods to detect specific antibody-antigen interactions are known in the art and can be used in the method, including, but not limited to, standard immunohistological methods, immunoprecipitation, an enzyme immunoassay, and a radioimmunoassay.

[00118] In general, the specific antibody will be detectably labeled, either directly or indirectly. Direct labels include radioisotopes; enzymes whose products are detectable (e.g., luciferase,  $\beta$ -galactosidase, and the like); fluorescent labels (e.g., fluorescein isothiocyanate, rhodamine, phycoerythrin, and the like); fluorescence emitting metals, e.g.,  $^{152}\text{Eu}$ , or others of the lanthanide series, attached to the antibody through metal chelating groups such as EDTA; chemiluminescent compounds, e.g., luminol, isoluminol, acridinium salts, and the like; bioluminescent compounds, e.g., luciferin, aequorin (green fluorescent protein), and the like.

[00119] The antibody may be attached (coupled) to an insoluble support, such as a polystyrene plate or a bead. Indirect labels include second antibodies specific for antibodies specific for the encoded polypeptide ("first specific antibody"), wherein the second antibody is labeled as described above; and members of specific binding pairs, e.g., biotin-avidin, and the like. The biological sample may be brought into contact with and immobilized on a solid support or carrier, such as nitrocellulose, that is capable of immobilizing cells, cell particles, or soluble proteins. The support may then be washed with suitable buffers, followed by contacting with a detectably-labeled first specific antibody. Detection methods are known in the art and will be chosen as appropriate to the signal emitted by the detectable label. Detection is generally accomplished in comparison to suitable controls, and to appropriate standards.

[00120] In some embodiments, the methods are adapted for use *in vivo*, e.g., to locate or identify sites where colon cancer cells are present. In these embodiments, a detectably-

labeled moiety, e.g., an antibody, which is specific for a colon cancer-associated polypeptide is administered to an individual (e.g., by injection), and labeled cells are located using standard imaging techniques, including, but not limited to, magnetic resonance imaging, computed tomography scanning, and the like. In this manner, colon cancer cells are differentially labeled.

Detecting a polynucleotide that represents a gene differentially expressed in a colon cancer cell

- [00121] In some embodiments, methods are provided for detecting a colon cancer cell by detecting expression in the cell of a transcript or that is differentially expressed in a colon cancer cell. Any of a variety of known methods can be used for detection, including, but not limited to, detection of a transcript by hybridization with a polynucleotide that hybridizes to a polynucleotide that is differentially expressed in a colon cancer cell; detection of a transcript by a polymerase chain reaction using specific oligonucleotide primers; *in situ* hybridization of a cell using as a probe a polynucleotide that hybridizes to a gene that is differentially expressed in a colon cancer cell.
- [00122] The methods can be used to detect and/or measure mRNA levels of a gene that is differentially expressed in a colon cancer cell. In some embodiments, the methods comprise: a) contacting a sample with a polynucleotide that corresponds to a differentially expressed gene described herein under conditions that allow hybridization; and b) detecting hybridization, if any. Detection of differential hybridization, when compared to a suitable control, is an indication of the presence in the sample of a polynucleotide that is differentially expressed in a colon cancer cell. Appropriate controls include, for example, a sample which is known not to contain a polynucleotide that is differentially expressed in a colon cancer cell, and use of a labeled polynucleotide of the same "sense" as the polynucleotide that is differentially expressed in a colon cancer cell. Conditions that allow hybridization are known in the art, and have been described in more detail above.
- [00123] Detection can also be accomplished by any known method, including, but not limited to, *in situ* hybridization, PCR (polymerase chain reaction), RT-PCR (reverse transcription-PCR), and "Northern" or RNA blotting, or combinations of such techniques,

using a suitably labeled polynucleotide. A variety of labels and labeling methods for polynucleotides are known in the art and can be used in the assay methods of the invention. Specific hybridization can be determined by comparison to appropriate controls.

[00124] Polynucleotide generally comprising at least 12 contiguous nt of a polynucleotide provided herein, as shown in the Sequence Listing or of the sequences of the genes corresponding to the polynucleotides of the Sequence Listing, are used for a variety of purposes, such as probes for detection of and/or measurement of, transcription levels of a polynucleotide that is differentially expressed in a colon cancer cell. Additional disclosure about preferred regions of the disclosed polynucleotide sequences is found in the Examples. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 2-, 5-, 10-, or 20-fold higher than the background hybridization provided with other unrelated sequences. It should be noted that "probe" as used in this context of detection of nucleic acid is meant to refer to a polynucleotide sequence used to detect a differentially expressed gene product in a test sample. As will be readily appreciated by the ordinarily skilled artisan, the probe can be detectably labeled and contacted with, for example, an array comprising immobilized polynucleotides obtained from a test sample (*e.g.*, mRNA). Alternatively, the probe can be immobilized on an array and the test sample detectably labeled. These and other variations of the methods of the invention are well within the skill in the art and are within the scope of the invention.

[00125] Nucleotide probes are used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots, mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization can be quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for *in situ* hybridization to cells to detect expression. Probes can also be used *in vivo* for diagnostic detection of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores,

fluorophores, and enzymes. Other examples of nucleotide hybridization assays are described in WO92/02526 and USPN 5,124,246.

[00126] PCR is another means for detecting small amounts of target nucleic acids (see, e.g., Mullis *et al.*, *Meth. Enzymol.* (1987) 155:335; USPN 4,683,195; and USPN 4,683,202). Two primer polynucleotides nucleotides that hybridize with the target nucleic acids are used to prime the reaction. The primers can be composed of sequence within or 3' and 5' to the polynucleotides of the Sequence Listing. Alternatively, if the primers are 3' and 5' to these polynucleotides, they need not hybridize to them or the complements. After amplification of the target with a thermostable polymerase, the amplified target nucleic acids can be detected by methods known in the art, e.g., Southern blot. mRNA or cDNA can also be detected by traditional blotting techniques (e.g., Southern blot, Northern blot, etc.) described in Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" (New York, Cold Spring Harbor Laboratory, 1989) (e.g., without PCR amplification). In general, mRNA or cDNA generated from mRNA using a polymerase enzyme can be purified and separated using gel electrophoresis, and transferred to a solid support, such as nitrocellulose. The solid support is exposed to a labeled probe, washed to remove any unhybridized probe, and duplexes containing the labeled probe are detected.

[00127] Methods using PCR amplification can be performed on the DNA from a single cell, although it is convenient to use at least about  $10^5$  cells. The use of the polymerase chain reaction is described in Saiki *et al.* (1985) *Science* 239:487, and a review of current techniques may be found in Sambrook, *et al.* Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2-14.33. A detectable label may be included in the amplification reaction. Suitable detectable labels include fluorochromes, (e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)), radioactive labels, (e.g.  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ , *etc.*), and the like. The label may be a two stage system, where the polynucleotides is conjugated to biotin, haptens, *etc.* having a high affinity binding partner,

*e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

#### Arrays

- [00128] Polynucleotide arrays provide a high throughput technique that can assay a large number of polynucleotides or polypeptides in a sample. This technology can be used as a tool to test for differential expression.
- [00129] A variety of methods of producing arrays, as well as variations of these methods, are known in the art and contemplated for use in the invention. For example, arrays can be created by spotting polynucleotide probes onto a substrate (*e.g.*, glass, nitrocellulose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions.
- [00130] Samples of polynucleotides can be detectably labeled (*e.g.*, using radioactive or fluorescent labels) and then hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Alternatively, the polynucleotides of the test sample can be immobilized on the array, and the probes detectably labeled. Techniques for constructing arrays and methods of using these arrays are described in, for example, Schena *et al.* (1996) *Proc Natl Acad Sci U S A.* 93(20):10614-9; Schena *et al.* (1995) *Science* 270(5235):467-70; Shalon *et al.* (1996) *Genome Res.* 6(7):639-45, USPN 5,807,522, EP 799 897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; USPN 5,593,839; USPN 5,578,832; EP 728 520; USPN 5,599,695; EP 721 016; USPN 5,556,752; WO 95/22058; and USPN 5,631,734.
- [00131] Arrays can be used to, for example, examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential expression of a gene corresponding to a polynucleotide described herein, where expression is compared between a test cell and control cell (*e.g.*, cancer cells and normal



cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific gene product. Exemplary uses of arrays are further described in, for example, Pappalarado *et al.*, *Sem. Radiation Oncol.* (1998) 8:217; and Ramsay, *Nature Biotechnol.* (1998) 16:40. Furthermore, many variations on methods of detection using arrays are well within the skill in the art and within the scope of the present invention. For example, rather than immobilizing the probe to a solid support, the test sample can be immobilized on a solid support which is then contacted with the probe.

Diagnosis, Prognosis, Assessment of Therapy (Therametrics), and Management of Cancer

- [00132] The polynucleotides described herein, as well as their gene products and corresponding genes and gene products, are of particular interest as genetic or biochemical markers (e.g., in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or to monitor the efficacy of various therapies and preventive interventions.
- [00133] For example, the level of expression of certain polynucleotides can be indicative of a poorer prognosis, and therefore warrant more aggressive chemo- or radio-therapy for a patient or vice versa. The correlation of novel surrogate tumor specific features with response to treatment and outcome in patients can define prognostic indicators that allow the design of tailored therapy based on the molecular profile of the tumor. These therapies include antibody targeting, antagonists (e.g., small molecules), and gene therapy.
- [00134] Determining expression of certain polynucleotides and comparison of a patient's profile with known expression in normal tissue and variants of the disease allows a determination of the best possible treatment for a patient, both in terms of specificity of treatment and in terms of comfort level of the patient. Surrogate tumor markers, such as polynucleotide expression, can also be used to better classify, and thus diagnose and treat, different forms and disease states of cancer. Two classifications widely used in oncology that can benefit from identification of the expression levels of the genes corresponding to

the polynucleotides described herein are staging of the cancerous disorder, and grading the nature of the cancerous tissue.

[00135] The polynucleotides that correspond to differentially expressed genes, as well as their encoded gene products, can be useful to monitor patients having or susceptible to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. In addition, the polynucleotides described herein, as well as the genes corresponding to such polynucleotides, can be useful as therametrics, *e.g.*, to assess the effectiveness of therapy by using the polynucleotides or their encoded gene products, to assess, for example, tumor burden in the patient before, during, and after therapy.

[00136] Furthermore, a polynucleotide identified as corresponding to a gene that is differentially expressed in, and thus is important for, one type of cancer can also have implications for development or risk of development of other types of cancer, *e.g.*, where a polynucleotide represents a gene differentially expressed across various cancer types. Thus, for example, expression of a polynucleotide corresponding to a gene that has clinical implications for metastatic colon cancer can also have clinical implications for breast cancer, prostate cancer, or ovarian cancer.

[00137] Staging. Staging is a process used by physicians to describe how advanced the cancerous state is in a patient. Staging assists the physician in determining a prognosis, planning treatment and evaluating the results of such treatment. Staging systems vary with the types of cancer, but generally involve the following "TNM" system: the type of tumor, indicated by T; whether the cancer has metastasized to nearby lymph nodes, indicated by N; and whether the cancer has metastasized to more distant parts of the body, indicated by M. Generally, if a cancer is only detectable in the area of the primary lesion without having spread to any lymph nodes, it is called Stage I or Stage II, depending on the degree of invasiveness as indicated by the tumor grade of the primary lesion. If the primary lesion is of tumor grade I or II and the patient does not have any regional or distant metastasis, the cancer is classified as Stage I. If the primary lesion is of tumor grade III or IV and the patient does not have any regional or distant metastasis, the cancer is classified as Stage II.

If the cancer has spread only to the regional lymph nodes, it is classified as Stage III. Cancers that have spread to a distant part of the body, such as liver, bone, brain or other sites, are Stage IV, the most advanced stage.

[00138] The polynucleotides and corresponding genes and gene products described herein can facilitate fine-tuning of the staging process by identifying markers for the aggressiveness of a cancer, *e.g.* the metastatic potential, as well as the presence in different areas of the body. Thus, a Stage II cancer with a polynucleotide signifying a high metastatic potential cancer can be used to change a borderline Stage II tumor to a Stage III tumor, justifying more aggressive therapy. Conversely, the presence of a polynucleotide signifying a lower metastatic potential allows more conservative staging of a tumor.

[00139] Grading of cancers. Grade is a term used to describe how closely a tumor resembles normal tissue of its same type. The microscopic appearance of a tumor is used to identify tumor grade based on parameters such as cell morphology, cellular organization, and other markers of differentiation. As a general rule, the grade of a tumor corresponds to its rate of growth or aggressiveness, with undifferentiated or high-grade tumors generally being more aggressive than well-differentiated or low-grade tumors. The following guidelines are generally used for grading tumors: 1) GX Grade cannot be assessed; 2) G1 Well differentiated; G2 Moderately well differentiated; 3) G3 Poorly differentiated; 4) G4 Undifferentiated.

[00140] The polynucleotides of the Sequence Listing, and their corresponding genes and gene products, can be especially valuable in determining the grade of the tumor, as they not only can aid in determining the differentiation status of the cells of a tumor, they can also identify factors other than differentiation that are valuable in determining the aggressiveness of a tumor, such as metastatic potential.

[00141] Assessment of proliferation of cells in tumor. The differential expression level of the polynucleotides described herein can facilitate assessment of the rate of proliferation of tumor cells, and thus provide an indicator of the aggressiveness of the rate of tumor growth. For example, assessment of the relative expression levels of genes involved in cell

cycle (e.g., TTK) can provide an indication of cellular proliferation, and thus serve as a marker of proliferation.

**[00142]**      Detection of colon cancer. The polynucleotides corresponding to genes that exhibit the appropriate expression pattern can be used to detect colon cancer in a subject. Colorectal cancer is one of the most common neoplasms in humans and perhaps the most frequent form of hereditary neoplasia. Prevention and early detection are key factors in controlling and curing colorectal cancer. Colorectal cancer begins as polyps, which are small, benign growths of cells that form on the inner lining of the colon. Over a period of several years, some of these polyps accumulate additional mutations and become cancerous. Multiple familial colorectal cancer disorders have been identified, which are summarized as follows: 1) Familial adenomatous polyposis (FAP); 2) Gardner's syndrome; 3) Hereditary nonpolyposis colon cancer (HNPCC); and 4) Familial colorectal cancer in Ashkenazi Jews.

**[00143]**      The expression of appropriate polynucleotides can be used in the diagnosis, prognosis and management of colorectal cancer. Detection of colon cancer can be determined using expression levels of any of these sequences alone or in combination with the levels of expression. Determination of the aggressive nature and/or the metastatic potential of a colon cancer can be determined by comparing levels of one or more gene products of the genes corresponding to the polynucleotides described herein, and comparing total levels of another sequence known to vary in cancerous tissue, e.g., expression of p53, DCC, ras, FAP (see, e.g., Fearon ER, *et al.*, *Cell* (1990) 61(5):759; Hamilton SR *et al.*, *Cancer* (1993) 72:957; Bodmer W, *et al.*, *Nat Genet.* (1994) 4(3):217; Fearon ER, *Ann N Y Acad Sci.* (1995) 768:101).

**[00144]**      For example, development of colon cancer can be detected by examining the level of expression of a gene corresponding to a polynucleotides described herein to the levels of oncogenes (e.g. ras) or tumor suppressor genes (e.g. FAP or p53). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous colon tissue, to discriminate between colon cancers with different cells of origin, to

discriminate between colon cancers with different potential metastatic rates, etc. For a review of markers of cancer, see, e.g., Hanahan et al. (2000) Cell 100:57-70.

Tumor classification and patient stratification

- [00145] The invention further provides for methods of classifying tumors, and thus grouping or "stratifying" patients, according to the expression profile of selected differentially expressed genes in a tumor. The expression patterns of differentially expressed genes can be analyzed for correlation with the expression patterns of other differentially expressed genes in a single tumor type or across tumor types. Genes that demonstrate consistent correlation can be grouped together, e.g., genes are grouped together where if one gene is overexpressed in a tumor, a second gene is also usually overexpressed. Tumors can then be classified according to the expression profile of one or more genes selected from one or more groups.
- [00146] For example, a colon tumor can be classified according to expression level of a gene product of one or more genes selected from one or more of the following groups: 1) Group I, which comprises the genes IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1; and 2) Group II, which comprises the genes IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1.
- [00147] A Group I-type colon tumor has increased expression of at least one, usually at least two, more usually at least three, even more usually at least four, preferably at least five, more preferably at least six or more, but usually not more than 12, 10, or 8, Group I genes relative to a non-cancerous colon cell, where the expression is increased at least about 1.5-fold, at least about 2-fold, at least about 5-fold, or at least about 10-fold, and can be as high 50-fold, but is usually not more than 20-fold or 30-fold.
- [00148] A Group II-type colon tumor is increased in expression of at least one, usually at least two, more usually at least three, Group II genes relative to a non-cancerous colon cells, where the expression is increased at least about 1.5-fold, at least about 2-fold, at least about 5-fold, or at least about 10-fold, and can be as high 50-fold, but is usually not more than 20-fold or 30-fold.

[00149] A Group I+II-type colon tumor is increased in expression of at least one, usually at least two, more usually at least three, even more usually at least four, preferably at least five, more preferably at least six or more, but usually not more than 12, 10, or 8, Group I genes relative to a non-cancerous colon cell, and has increased expression of at least one, usually at least two, more usually at least three, Group II genes relative to a non-cancerous colon cells, where expression of both the Group I and Group II genes is increased at least about 1.5-fold, at least about 2-fold, at least about 5-fold, or at least about 10-fold, and can be as high 50-fold, but is usually not more than 20-fold or 30-fold.

[00150] The tumor of each patient in a pool of potential patients for a clinical trial can be classified as described above. Patients having similarly classified tumors can then be selected for participation in an investigative or clinical trial of a cancer therapeutic where a homogeneous population is desired. The tumor classification of a patient can also be used in assessing the efficacy of a cancer therapeutic in a heterogeneous patient population. Thus, comparison of an individual's expression profile to the population profile for a type of cancer, permits the selection or design of drugs or other therapeutic regimens that are expected to be safe and efficacious for a particular patient or patient population (i.e., a group of patients having the same type of cancer).

[00151] In addition, the ability to target populations expected to show the most clinical benefit, based on expression profile can enable: 1) the repositioning of already marketed drugs; 2) the rescue of drug candidates whose clinical development has been discontinued as a result of safety or efficacy limitations, which are patient subgroup-specific; and 3) an accelerated and less costly development for candidate therapeutics and more optimal drug labeling (e.g. since measuring the effect of various doses of an agent on patients with a particular expression profile is useful for optimizing effective dose).

#### Treatment of colon cancer

[00152] The invention further provides methods for reducing growth of colon cancer cells. In general, the methods comprise contacting a colon cancer cell with a substance that modulates (1) expression of a polynucleotide corresponding to a gene that is differentially expressed in colon cancer; or (2) a level of and/or an activity of a colon cancer-associated

polypeptide. In general, the methods provide for decreasing the expression of a gene that is differentially expressed in a colon cancer cell or decreasing the level of and/or decreasing an activity of a colon cancer-associated polypeptide.

**[00153]** "Reducing growth of colon cancer cells" includes, but is not limited to, reducing proliferation of colon cancer cells relative to untreated colon cancer cells, as well as reducing the incidence of a non-cancerous colon cell becoming a cancerous colon cell relative to untreated colon cancer cells. Whether a reduction in colon cancer cell growth has been achieved can be readily determined using any known assay, including, but not limited to, [<sup>3</sup>H]-thymidine incorporation; counting cell number over a period of time; detecting and/or measuring a marker associated with colon cancer (e.g., CEA, CA19-9, and LASA).

**[00154]** The present invention provides methods for treating colon cancer, generally comprising administering to an individual in need thereof a substance that reduces colon cancer cell growth, in an amount sufficient to reduce colon cancer cell growth and treat the colon cancer. Whether a substance, or a specific amount of the substance, is effective in treating colon cancer can be assessed using any of a variety of known diagnostic assays for colon cancer, including, but not limited to, sigmoidoscopy, proctoscopy, rectal examination, colonoscopy with biopsy, contrast radiographic studies, CAT scans, angiography, and detection of a tumor marker associated with colon cancer in the blood of the individual. The substance can be administered systemically or locally. Thus, in some embodiments, the substance is administered locally, and colon cancer growth is decreased at the site of administration. Local administration may be useful in treating, e.g., a solid tumor.

**[00155]** A substance that reduces colon cancer cell growth can be targeted to a colon cancer cell. Thus, in some embodiments, the invention provides a method of delivering a drug to a colon cancer cell, comprising administering a drug-antibody complex to a subject, wherein the antibody is specific for a colon cancer-associated polypeptide, and the drug is one that reduces colon cancer cell growth, a variety of which are known in the art. Targeting can be accomplished by coupling (e.g., linking, directly or via a linker molecule,

either covalently or non-covalently, so as to form a drug-antibody complex) a drug to an antibody specific for a colon cancer-associated polypeptide. Methods of coupling a drug to an antibody are well known in the art and need not be elaborated upon herein.

**[00156]** In another embodiment, differentially expressed gene products (*e.g.*, polypeptides or polynucleotides encoding such polypeptides) may be effectively used in treatment through vaccination. The growth of cancer cells is naturally limited in part due to immune surveillance. Stimulation of the immune system using a particular tumor-specific antigen enhances the effect towards the tumor expressing the antigen. An active vaccine comprising a polypeptide encoded by the cDNA of this invention would be appropriately administered to subjects having overabundance of the corresponding RNA, or those predisposed for developing cancer cells with overabundance of the same RNA. Polypeptide antigens are typically combined with an adjuvant as part of a vaccine composition. The vaccine is preferably administered first as a priming dose, and then again as a boosting dose, usually at least four weeks later. Further boosting doses may be given to enhance the effect. The dose and its timing are usually determined by the person responsible for the treatment.

**[00157]** The invention also encompasses the selection of a therapeutic regimen is then based upon the expression profile of differentially expressed genes in the patient's tumor. For example, a tumor can be analyzed for the expression pattern of a gene product of one or more genes selected from one or more of the following groups: 1) Group I, which comprises the genes IGF2, TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1; and 2) Group II, which comprises the genes IFITM (1-8U; 1-8D; 9-27), ITAK, and BIRC3/H-IAP1. The tumor can then be classified as a Group I-type, Group II-type, or Group I+II-type tumor based on the expression profile of the tumor. The expression patterns are then compared to the expression patterns of tumors that are known to respond to a selected therapy. Where the expression profiles of the test tumor cell and the expression profile of a tumor cell of known drug responsivity at least substantially match (*e.g.*, both are members of the same classification group (*e.g.*, Group I, Group II, Group I+II), then the drug selected for



therapy is the drug to which tumors with that expression pattern respond. For example, if a patient's tumor indicates that the tumor can be classified as a "Group I" tumor, then a drug having efficacy in the treatment of such tumors is selected for therapy of that patient.

#### Identification of Therapeutic Targets and Anti-Cancer Therapeutic Agents

[00158] The present invention also encompasses methods for identification of agents having the ability to modulate activity of a differentially expressed gene product, as well as methods for identifying a differentially expressed gene product as a therapeutic target for treatment of cancer, especially colon cancer.

#### Candidate agents

[00159] Identification of compounds that modulate activity of a differentially expressed gene product can be accomplished using any of a variety of drug screening techniques. Such agents are candidates for development of cancer therapies. Of particular interest are screening assays for agents that have tolerable toxicity for normal, non-cancerous human cells. The screening assays of the invention are generally based upon the ability of the agent to modulate an activity of a differentially expressed gene product and/or to inhibit or suppress phenomenon associated with cancer (*e.g.*, cell proliferation, colony formation, cell cycle arrest, metastasis, and the like).

[00160] The term "agent" as used herein describes any molecule, *e.g.* protein or pharmaceutical, with the capability of modulating a biological activity of a gene product of a differentially expressed gene. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.* at zero concentration or below the level of detection.

[00161] Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at

least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including, but not limited to: peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

[00162] Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts (including extracts from human tissue to identify endogenous factors affecting differentially expressed gene products) are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

[00163] Exemplary candidate agents of particular interest include, but are not limited to, antisense polynucleotides, and antibodies, soluble receptors, and the like. Antibodies and soluble receptors are of particular interest as candidate agents where the target differentially expressed gene product is secreted or accessible at the cell-surface (*e.g.*, receptors and other molecule stably-associated with the outer cell membrane).

#### Screening of candidate agents

[00164] Screening assays can be based upon any of a variety of techniques readily available and known to one of ordinary skill in the art. In general, the screening assays involve contacting a cancerous cell (preferably a cancerous colon cell) with a candidate agent, and assessing the effect upon biological activity of a differentially expressed gene product. The effect upon a biological activity can be detected by, for example, detection of expression of a gene product of a differentially expressed gene (*e.g.*, a decrease in mRNA

or polypeptide levels, would in turn cause a decrease in biological activity of the gene product). Alternatively or in addition, the effect of the candidate agent can be assessed by examining the effect of the candidate agent in a functional assay. For example, where the differentially expressed gene product is an enzyme, then the effect upon biological activity can be assessed by detecting a level of enzymatic activity associated with the differentially expressed gene product. The functional assay will be selected according to the differentially expressed gene product. In general, where the differentially expressed gene is increased in expression in a cancerous cell, agents of interest are those that decrease activity of the differentially expressed gene product.

[00165] Assays described *infra* can be readily adapted in the screening assay embodiments of the invention. Exemplary assays useful in screening candidate agents include, but are not limited to, hybridization-based assays (*e.g.*, use of nucleic acid probes or primers to assess expression levels), antibody-based assays (*e.g.*, to assess levels of polypeptide gene products), binding assays (*e.g.*, to detect interaction of a candidate agent with a differentially expressed polypeptide, which assays may be competitive assays where a natural or synthetic ligand for the polypeptide is available), and the like. Additional exemplary assays include, but are not necessarily limited to, cell proliferation assays, antisense knockout assays, assays to detect inhibition of cell cycle, assays of induction of cell death/apoptosis, and the like. Generally such assays are conducted *in vitro*, but many assays can be adapted for *in vivo* analyses, *e.g.*, in an animal model of the cancer.

#### Identification of therapeutic targets

[00166] In another embodiment, the invention contemplates identification of differentially expressed genes and gene products as therapeutic targets. In some respects, this is the converse of the assays described above for identification of agents having activity in modulating (*e.g.*, decreasing or increasing) activity of a differentially expressed gene product.

[00167] In this embodiment, therapeutic targets are identified by examining the effect(s) of an agent that can be demonstrated or has been demonstrated to modulate a cancerous phenotype (*e.g.*, inhibit or suppress or prevent development of a cancerous phenotype).

Such agents are generally referred to herein as an "anti-cancer agent", which agents encompass chemotherapeutic agents. For example, the agent can be an antisense oligonucleotide that is specific for a selected gene transcript. For example, the antisense oligonucleotide may have a sequence corresponding to a sequence of a differentially expressed gene described herein, *e.g.*, a sequence of one of SEQ ID NOS:1-20.

[00168] Assays for identification of therapeutic targets can be conducted in a variety of ways using methods that are well known to one of ordinary skill in the art. For example, a test cancerous cell that expresses or overexpresses a differentially expressed gene is contacted with an anti-cancer agent, the effect upon a cancerous phenotype and a biological activity of the candidate gene product assessed. The biological activity of the candidate gene product can be assayed by examining, for example, modulation of expression of a gene encoding the candidate gene product (*e.g.*, as detected by, for example, an increase or decrease in transcript levels or polypeptide levels), or modulation of an enzymatic or other activity of the gene product. The cancerous phenotype can be, for example, cellular proliferation, loss of contact inhibition of growth (*e.g.*, colony formation), tumor growth (*in vitro* or *in vivo*), and the like. Alternatively or in addition, the effect of modulation of a biological activity of the candidate target gene upon cell death/apoptosis or cell cycle regulation can be assessed.

[00169] Inhibition or suppression of a cancerous phenotype, or an increase in cell/death apoptosis as a result of modulation of biological activity of a candidate gene product indicates that the candidate gene product is a suitable target for cancer therapy. Assays described *infra* can be readily adapted in for assays for identification of therapeutic targets. Generally such assays are conducted *in vitro*, but many assays can be adapted for *in vivo* analyses, *e.g.*, in an appropriate, art-accepted animal model of the cancer.

#### Use of Polypeptides to Screen for Peptide Analogs and Antagonists

[00170] Polypeptides encoded by differentially expressed genes identified herein can be used to screen peptide libraries to identify binding partners, such as receptors, from among

the encoded polypeptides. Peptide libraries can be synthesized according to methods known in the art (*see, e.g.*, USPN 5,010,175 and WO 91/17823).

[00171] Agonists or antagonists of the polypeptides if the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, etc. The assay conditions ideally should resemble the conditions under which the native activity is exhibited *in vivo*, that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

[00172] Such screening and experimentation can lead to identification of a polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide described herein, and at least one peptide agonist or antagonist of the binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

#### Vaccines and Uses

[00173] The differentially expressed nucleic acids and polypeptides produced by the nucleic acids of the invention can also be used to modulate primary immune response to prevent or treat cancer. Every immune response is a complex and intricately regulated sequence of events involving several cell types. It is triggered when an antigen enters the body and encounters a specialized class of cells called antigen-presenting cells (APCs). These APCs capture a minute amount of the antigen and display it in a form that can be recognized by

antigen-specific helper T lymphocytes. The helper (Th) cells become activated and, in turn, promote the activation of other classes of lymphocytes, such as B cells or cytotoxic T cells. The activated lymphocytes then proliferate and carry out their specific effector functions, which in many cases successfully activate or eliminate the antigen. Thus, activating the immune response to a particular antigen associated with a cancer cell can protect the patient from developing cancer or result in lymphocytes eliminating cancer cells expressing the antigen.

[00174] Gene products, including polypeptides, mRNA (particularly mRNAs having distinct secondary and/or tertiary structures), cDNA, or complete gene, can be prepared and used in vaccines for the treatment or prevention of hyperproliferative disorders and cancers. The nucleic acids and polypeptides can be utilized to enhance the immune response, prevent tumor progression, prevent hyperproliferative cell growth, and the like. Methods for selecting nucleic acids and polypeptides that are capable of enhancing the immune response are known in the art. Preferably, the gene products for use in a vaccine are gene products which are present on the surface of a cell and are recognizable by lymphocytes and antibodies.

[00175] The gene products may be formulated with pharmaceutically acceptable carriers into pharmaceutical compositions by methods known in the art. The composition is useful as a vaccine to prevent or treat cancer. The composition may further comprise at least one co-immunostimulatory molecule, including but not limited to one or more major histocompatibility complex (MHC) molecules, such as a class I or class II molecule, preferably a class I molecule. The composition may further comprise other stimulator molecules including B7.1, B7.2, ICAM-1, ICAM-2, LFA-1, LFA-3, CD72 and the like, immunostimulatory polynucleotides (which comprise an 5'-CG-3' wherein the cytosine is unmethylated), and cytokines which include but are not limited to IL-1 through IL-15, TNF- $\alpha$ , IFN- $\gamma$ , RANTES, G-CSF, M-CSF, IFN- $\alpha$ , CTAP III, ENA-78, GRO, I-309, PF-4, IP-10, LD-78, MGSA, MIP-1 $\alpha$ , MIP-1 $\beta$ , or combination thereof, and the like for immunopotential. In one embodiment, the immunopotential of particular interest are those which facilitate a Th1 immune response.

- [00176] The gene products may also be prepared with a carrier that will protect the gene products against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, polylactic acid, and the like. Methods for preparation of such formulations are known in the art.
- [00177] In the methods of preventing or treating cancer, the gene products may be administered via one of several routes including but not limited to transdermal, transmucosal, intravenous, intramuscular, subcutaneous, intradermal, intraperitoneal, intrathecal, intrapleural, intrauterine, rectal, vaginal, topical, intratumor, and the like. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be by nasal sprays or suppositories. For oral administration, the gene products are formulated into conventional oral administration form such as capsules, tablets and toxics.
- [00178] The gene product is administered to a patient in an amount effective to prevent or treat cancer. In general, it is desirable to provide the patient with a dosage of gene product of at least about 1 pg per Kg body weight, preferably at least about 1 ng per Kg body weight, more preferably at least about 1  $\mu$ g or greater per Kg body weight of the recipient. A range of from about 1 ng per Kg body weight to about 100 mg per Kg body weight is preferred although a lower or higher dose may be administered. The dose is effective to prime, stimulate and/or cause the clonal expansion of antigen-specific T lymphocytes, preferably cytotoxic T lymphocytes, which in turn are capable of preventing or treating cancer in the recipient. The dose is administered at least once and may be provided as a bolus or a continuous administration. Multiple administrations of the dose over a period of several weeks to months may be preferable. Subsequent doses may be administered as indicated.

[00179] In another method of treatment, autologous cytotoxic lymphocytes or tumor infiltrating lymphocytes may be obtained from a patient with cancer. The lymphocytes are grown in culture, and antigen-specific lymphocytes are expanded by culturing in the presence of the specific gene products alone or in combination with at least one co-immunostimulatory molecule with cytokines. The antigen-specific lymphocytes are then infused back into the patient in an amount effective to reduce or eliminate the tumors in the patient. Cancer vaccines and their uses are further described in USPN 5,961,978; USPN 5,993,829; USPN 6,132,980; and WO 00/38706.

Pharmaceutical Compositions and Uses

[00180] Pharmaceutical compositions can comprise polypeptides, receptors that specifically bind a polypeptide produced by a differentially expressed gene (*e.g.*, antibodies, or polynucleotides (including antisense nucleotides and ribozymes) of the claimed invention in a therapeutically effective amount. The compositions can be used to treat primary tumors as well as metastases of primary tumors. In addition, the pharmaceutical compositions can be used in conjunction with conventional methods of cancer treatment, *e.g.*, to sensitize tumors to radiation or conventional chemotherapy.

[00181] Where the pharmaceutical composition comprises a receptor (such as an antibody) that specifically binds to a gene product encoded by a differentially expressed gene, the receptor can be coupled to a drug for delivery to a treatment site or coupled to a detectable label to facilitate imaging of a site comprising colon cancer cells. Methods for coupling antibodies to drugs and detectable labels are well known in the art, as are methods for imaging using detectable labels.

[00182] The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature.



[00183] The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

[00184] A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles.

[00185] Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, e.g., mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in *Remington: The*

*Science and Practice of Pharmacy* (1995) Alfonso Gennaro, Lippincott, Williams, & Wilkins.

Delivery Methods

- [00186] Once formulated, the compositions contemplated by the invention can be (1) administered directly to the subject (e.g., as polynucleotide, polypeptides, small molecule agonists or antagonists, and the like); or (2) delivered *ex vivo*, to cells derived from the subject (e.g., as in *ex vivo* gene therapy). Direct delivery of the compositions will generally be accomplished by parenteral injection, e.g., subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumoral or to the interstitial space of a tissue. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.
- [00187] Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in e.g., International Publication No. WO 93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.
- [00188] Once differential expression of a gene corresponding to a polynucleotide described herein has been found to correlate with a proliferative disorder, such as neoplasia, dysplasia, and hyperplasia, the disorder can be amenable to treatment by administration of a therapeutic agent based on the provided polynucleotide, corresponding polypeptide or other corresponding molecule (e.g., antisense, ribozyme, etc.). In other embodiments, the disorder can be amenable to treatment by administration of a small molecule drug that, for example, serves as an inhibitor (antagonist) of the function of the encoded gene product of

a gene having increased expression in cancerous cells relative to normal cells or as an agonist for gene products that are decreased in expression in cancerous cells (*e.g.*, to promote the activity of gene products that act as tumor suppressors).

**[00189]** The dose and the means of administration of the inventive pharmaceutical compositions are determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. For example, administration of polynucleotide therapeutic composition agents includes local or systemic administration, including injection, oral administration, particle gun or catheterized administration, and topical administration. In general, the therapeutic polynucleotide composition contains an expression construct comprising a promoter operably linked to a polynucleotide of at least 12, 22, 25, 30, or 35 contiguous nt of the polynucleotide disclosed herein. Various methods can be used to administer the therapeutic composition directly to a specific site in the body. For example, a small metastatic lesion is located and the therapeutic composition injected several times in several different locations within the body of tumor. Alternatively, arteries which serve a tumor are identified, and the therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor. A tumor that has a necrotic center is aspirated and the composition injected directly into the now empty center of the tumor. The antisense composition is directly administered to the surface of the tumor, for example, by topical application of the composition. X-ray imaging is used to assist in certain of the above delivery methods.

**[00190]** Targeted delivery of therapeutic compositions containing an antisense polynucleotide, subgenomic polynucleotides, or antibodies to specific tissues can also be used. Receptor-mediated DNA delivery techniques are described in, for example, Findeis *et al.*, *Trends Biotechnol.* (1993) 11:202; Chiou *et al.*, *Gene Therapeutics: Methods And Applications Of Direct Gene Transfer* (J.A. Wolff, ed.) (1994); Wu *et al.*, *J. Biol. Chem.* (1988) 263:621; Wu *et al.*, *J. Biol. Chem.* (1994) 269:542; Zenke *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1990) 87:3655; Wu *et al.*, *J. Biol. Chem.* (1991) 266:338. Therapeutic compositions containing a polynucleotide are administered in a range of about 100 ng to

about 200 mg of DNA for local administration in a gene therapy protocol. Concentration ranges of about 500 ng to about 50 mg, about 1  $\mu$ g to about 2 mg, about 5  $\mu$ g to about 500  $\mu$ g, and about 20  $\mu$ g to about 100  $\mu$ g of DNA can also be used during a gene therapy protocol. Factors such as method of action (e.g., for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations which will affect the dosage required for ultimate efficacy of the antisense subgenomic polynucleotides.

[00191] Where greater expression is desired over a larger area of tissue, larger amounts of antisense subgenomic polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of, for example, a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect. For polynucleotide related genes encoding polypeptides or proteins with anti-inflammatory activity, suitable use, doses, and administration are described in USPN 5,654,173.

[00192] The therapeutic polynucleotides and polypeptides of the present invention can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally, Jolly, *Cancer Gene Therapy* (1994) 1:51; Kimura, *Human Gene Therapy* (1994) 5:845; Connelly, *Human Gene Therapy* (1995) 1:185; and Kaplitt, *Nature Genetics* (1994) 6:148). Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

[00193] Viral-based vectors for delivery of a desired polynucleotide and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (see, e.g., WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; USPN 5, 219,740; WO 93/11230; WO 93/10218; USPN 4,777,127; GB Patent No. 2,200,651; EP 0 345 242; and WO 91/02805), alphavirus-based vectors (e.g., Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis

virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), and adeno-associated virus (AAV) vectors (see, e.g., WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655). Administration of DNA linked to killed adenovirus as described in Curiel, *Hum. Gene Ther.* (1992) 3:147 can also be employed.

**[00194]** Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone (see, e.g., Curiel, *Hum. Gene Ther.* (1992) 3:147); ligand-linked DNA (see, e.g., Wu, *J. Biol. Chem.* (1989) 264:16985); eukaryotic cell delivery vehicles cells (see, e.g., USPN 5,814,482; WO 95/07994; WO 96/17072; WO 95/30763; and WO 97/42338) and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and USPN 5,580,859. Liposomes that can act as gene delivery vehicles are described in USPN 5,422,120; WO 95/13796; WO 94/23697; WO 91/14445; and EP 0524968. Additional approaches are described in Philip, *Mol. Cell Biol.* (1994) 14:2411, and in Woffendin, *Proc. Natl. Acad. Sci.* (1994) 91:1581.

**[00195]** Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al.*, *Proc. Natl. Acad. Sci. USA* (1994) 91(24):11581. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials or use of ionizing radiation (see, e.g., USPN 5,206,152 and WO 92/11033). Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun (see, e.g., USPN 5,149,655); use of ionizing radiation for activating transferred gene (see, e.g., USPN 5,206,152 and WO 92/11033).

#### EXAMPLES

**[00196]** The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only

experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

- [00197] The Examples below demonstrate the successful use of strategies of the invention to identify genes that are differentially expressed in colon tumor cells relative to normal colon cells, and further how these genes can serve as the basis for classification of tumor types and the stratification of patient shaving such tumors. These differentially expressed genes, which are represented by the polynucleotides referred to herein by SEQ ID NOS:1-20, are expressed at a higher or lower levels in colon tumor cells relative to their expression in normal cells of same tissue.

#### EXAMPLE 1: SOURCE OF BIOLOGICAL MATERIALS

- [00198] The biological materials used in the experiments that led to the present invention are described below.

##### Source of patient tissue samples

- [00199] Normal and cancerous tissues were collected from patients using laser capture microdissection (LCM) techniques, which techniques are well known in the art (see, e.g., Ohyama *et al.* (2000) *Biotechniques* 29:530-6; Curran *et al.* (2000) *Mol. Pathol.* 53:64-8; Suarez-Quian *et al.* (1999) *Biotechniques* 26:328-35; Simone *et al.* (1998) *Trends Genet* 14:272-6; Conia *et al.* (1997) *J. Clin. Lab. Anal.* 11:28-38; Emmert-Buck *et al.* (1996) *Science* 274:998-1001). Table 1 (inserted prior to claims) provides information about each patient from which colon tissue samples were isolated, including: the Patient ID ("PT ID") and Path ReportID ("Path ID"), which are numbers assigned to the patient and the pathology reports for identification purposes; the group ("Grp") to which the patients have been assigned; the anatomical location of the tumor ("Anatom Loc"); the primary tumor size ("Size"); the primary tumor grade ("Grade"); the identification of the histopathological grade ("Histo Grade"); a description of local sites to which the tumor had

invaded ("Local Invasion"); the presence of lymph node metastases ("Lymph Met"); the incidence of lymph node metastases (provided as a number of lymph nodes positive for metastasis over the number of lymph nodes examined) ("Lymph Met Incid"); the regional lymphnode grade ("Reg Lymph Grade"); the identification or detection of metastases to sites distant to the tumor and their location ("Dist Met & Loc"); the grade of distant metastasis ("Dist Met Grade"); and general comments about the patient or the tumor ("Comments"). Histopathology of all primary tumors indicated the tumor was adenocarcinoma except for Patient ID Nos. 130 (for which no information was provided), 392 (in which greater than 50% of the cells were mucinous carcinoma), and 784 (adenosquamous carcinoma). Extranodal extensions were described in three patients, Patient ID Nos. 784 and 791. Lymphovascular invasion was described in Patient ID Nos. 128, 228, 278, 517, 784, 786, 791, and 890. Crohn's-like infiltrates were described in seven patients, Patient ID Nos. 52, 264, 268, 392, 393, 784, and 791.

Source of polynucleotides on arrays

[00200] Polynucleotides for use on the arrays were obtained from both publicly available sources and from cDNA libraries generated from selected cell lines and patient tissues. Table 2 (inserted prior to claims) provides information about the polynucleotides on the arrays including: (1) the "SEQ ID NO" assigned to each sequence for use in the present specification; (2) the spot identification number ("Spot ID"), an internal reference that serves as a unique identifier for the spot on the array; (3) the "Clone ID" assigned to the clone from which the sequence was isolated; (4) the number of the Group ("Grp") to which the gene is assigned (see Example 3 below); and (5) the gene represented by the SEQ ID NO ("Gene").

[00201] The sequences corresponding to the SEQ ID NOS are provided in the Sequence Listing.

Characterization of sequences

[00202] The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the RepeatMasker masking program, publicly available through a web site supported by the University of Washington (*See also* Smit, A.F.A. and

Green, P., unpublished results). Generally, masking does not influence the final search results, except to eliminate sequences of relatively little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats. Masking resulted in the elimination of several sequences.

[00203] The remaining sequences of the isolated polynucleotides were used in a homology search of the GenBank database using the TeraBLAST program (TimeLogic, Crystal Bay, Nevada), a DNA and protein sequence homology searching algorithm. TeraBLAST is a version of the publicly available BLAST search algorithm developed by the National Center for Biotechnology, modified to operate at an accelerated speed with increased sensitivity on a specialized computer hardware platform. The program was run with the default parameters recommended by TimeLogic to provide the best sensitivity and speed for searching DNA and protein sequences. Gene assignment for the query sequences was determined based on best hit from the GenBank database; expectancy values are provided with the hit.

#### Summary of TeraBLAST Search Results

[00204] Table 2 also provides information about the gene corresponding to each polynucleotide. Table 2 includes: (1) the "SEQ ID NO" of the sequence; (2) the GenBank Accession Number of the publicly available sequence corresponding to the polynucleotide ("GBHit"); (3) a description of the GenBank sequence ("GBDesc"); (4) the score of the similarity of the polynucleotide sequence and the GenBank sequence ("GBScore"). The published information for each GenBank and EST description, as well as the corresponding sequence identified by the provided accession number, are incorporated herein by reference.

#### EXAMPLE 2: DETECTION OF DIFFERENTIAL EXPRESSION USING ARRAYS

[00205] cDNA probes were prepared from total RNA isolated from the patient cells described above. Since LCM provides for the isolation of specific cell types to provide a substantially homogenous cell sample, this provided for a similarly pure RNA sample.



- [00206] Total RNA was first reverse transcribed into cDNA using a primer containing a T7 RNA polymerase promoter, followed by second strand DNA synthesis. cDNA was then transcribed *in vitro* to produce antisense RNA using the T7 promoter-mediated expression (see, e.g., Luo *et al.* (1999) *Nature Med* 5:117-122), and the antisense RNA was then converted into cDNA. The second set of cDNAs were again transcribed *in vitro*, using the T7 promoter, to provide antisense RNA. Optionally, the RNA was again converted into cDNA, allowing for up to a third round of T7-mediated amplification to produce more antisense RNA. Thus the procedure provided for two or three rounds of *in vitro* transcription to produce the final RNA used for fluorescent labeling.
- [00207] Fluorescent probes were generated by first adding control RNA to the antisense RNA mix, and producing fluorescently labeled cDNA from the RNA starting material. Fluorescently labeled cDNAs prepared from the tumor RNA sample were compared to fluorescently labeled cDNAs prepared from normal cell RNA sample. For example, the cDNA probes from the normal cells were labeled with Cy3 fluorescent dye (green) and the cDNA probes prepared from the tumor cells were labeled with Cy5 fluorescent dye (red), and vice versa.
- [00208] Each array used had an identical spatial layout and control spot set. Each microarray was divided into two areas, each area having an array with, on each half, twelve groupings of 32 x 12 spots, for a total of about 9,216 spots on each array. The two areas are spotted identically which provide for at least two duplicates of each clone per array.
- [00209] Polynucleotides for use on the arrays were obtained from both publicly available sources and from cDNA libraries generated from selected cell lines and patient tissues as described above and in Table 2. PCR products of from about 0.5kb to 2.0 kb amplified from these sources were spotted onto the array using a Molecular Dynamics Gen III spotter according to the manufacturer's recommendations. The first row of each of the 24 regions on the array had about 32 control spots, including 4 negative control spots and 8 test polynucleotides. The test polynucleotides were spiked into each sample before the labeling reaction with a range of concentrations from 2-600 pg/slide and ratios of 1:1. For

each array design, two slides were hybridized with the test samples reverse-labeled in the labeling reaction. This provided for about four duplicate measurements for each clone, two of one color and two of the other, for each sample.

**[00210]** The differential expression assay was performed by mixing equal amounts of probes from tumor cells and normal cells of the same patient. The arrays were prehybridized by incubation for about 2 hrs at 60°C in 5X SSC/0.2% SDS/1 mM EDTA, and then washed three times in water and twice in isopropanol. Following prehybridization of the array, the probe mixture was then hybridized to the array under conditions of high stringency (overnight at 42°C in 50% formamide, 5X SSC, and 0.2% SDS. After hybridization, the array was washed at 55°C three times as follows: 1) first wash in 1X SSC/0.2% SDS; 2) second wash in 0.1X SSC/0.2% SDS; and 3) third wash in 0.1X SSC.

**[00211]** The arrays were then scanned for green and red fluorescence using a Molecular Dynamics Generation III dual color laser-scanner/detector. The images were processed using BioDiscovery Autogene software, and the data from each scan set normalized to provide for a ratio of expression relative to normal. Data from the microarray experiments was analyzed according to the algorithms described in U.S. application serial no. 60/252,358, filed November 20, 2000, by E.J. Moler, M.A. Boyle, and F.M. Randazzo, and entitled "Precision and accuracy in cDNA microarray data," which application is specifically incorporated herein by reference.

**[00212]** The experiment was repeated, this time labeling the two probes with the opposite color in order to perform the assay in both "color directions." Each experiment was sometimes repeated with two more slides (one in each color direction). The level fluorescence for each sequence on the array expressed as a ratio of the geometric mean of 8 replicate spots/genes from the four arrays or 4 replicate spots/gene from 2 arrays or some other permutation. The data were normalized using the spiked positive controls present in each duplicated area, and the precision of this normalization was included in the final determination of the significance of each differential. The fluorescent intensity of each

spot was also compared to the negative controls in each duplicated area to determine which spots have detected significant expression levels in each sample.

[00213] A statistical analysis of the fluorescent intensities was applied to each set of duplicate spots to assess the precision and significance of each differential measurement, resulting in a p-value testing the null hypothesis that there is no differential in the expression level between the tumor and normal samples of each patient. During initial analysis of the microarrays, the hypothesis was accepted if  $p > 10^{-3}$ , and the differential ratio was set to 1.000 for those spots. All other spots have a significant difference in expression between the tumor and normal sample. If the tumor sample has detectable expression and the normal does not, the ratio is truncated at 1000 since the value for expression in the normal sample would be zero, and the ratio would not be a mathematically useful value (e.g., infinity). If the normal sample has detectable expression and the tumor does not, the ratio is truncated to 0.001, since the value for expression in the tumor sample would be zero and the ratio would not be a mathematically useful value. These latter two situations are referred to herein as "on/off." Database tables were populated using a 95% confidence level ( $p > 0.05$ ).

[00214] Table 3 summarizes the results of the differential expression analysis. Table 3 provides: (1) the spot identification number ("Spot ID"), an internal reference that serves as a unique identifier for the spot on the array; (2) the number of the Group ("Grp") to which the gene is assigned (see Example 3 below); and (3) the ratio of expression of the gene in each of the patient samples, identified by the patient ID number (e.g., 15). This data represents the ratio of differential expression for the samples tested from that particular patient's tissues (e.g., "RATIO15" is the ratio from the tissue samples of Patient ID no. 15). The ratios of differential expression are expressed as a normalized hybridization signal associated with the tumor probe divided by the normalized hybridization signal with the normal probe. Thus, a ratio greater than 1 indicates that the gene product is increased in expression in cancerous cells relative to normal cells, while a ratio of less than 1 indicates the opposite.

[00215] These data provide evidence that the genes represented by the polynucleotides having the indicated sequences are differentially expressed in colon cancer as compared to normal non-cancerous colon tissue.

EXAMPLE 3: STRATIFICATION OF COLON CANCERS USING DIFFERENTIAL EXPRESSION DATA

[00216] Groups of genes with differential expression data correlating with specific genes of interest can be identified using statistical analysis such as the Student t-test and Spearman rank correlation (Stanton Glantz (1997) Primer of Bio-Statistics, McGraw Hill, pp 65-107, 256-262). Using these statistical tests, patients having tumors that exhibit similar differential expression patterns can be assigned to Groups. At least two Groups were identified, and are described below.

Group 1

Genes that exhibit differential expression in colon cancer in a pattern that correlates with IGF2

[00217] Using both the Student-t test and the Spearman rank correlation test, the differential expression data of IGF2 correlated with that of 14 distinct genes: TTK, MAPKAPK2, MARCKS, BBS2, CETN2 CGI-148 protein, FGFR4, FHL3, FLJ22066, KIP2, MGC:29604, NQO2, and OGG1 (see Table 3). The differential expression data for these genes is presented in graphical form in Figs. 1-16. This group was identified as Group 1. IGF2 is a secreted protein and has been reported to be involved in colon as well as other cancers (Toretsky JA and Helman LJ (1996) J Endocrinol 149(3):367-72). Genes whose expression patterns correlate with IGF2 may provide a mechanism for the involvement of IGF2 in cancer. Among the genes in Group 1 are genes such as TTK (a kinase implicated in mitotic spindle check point), MAP-KAP kinase 2 (mitogen-activated protein (MAP) kinase activated protein kinase 2), and MARCKS (myristoylated alanine-rich C kinase substrate, which is a substrate of protein kinase C). The protein products of these genes and their associated signaling pathways can be targets for small molecule drug development for anti-cancer therapy. Furthermore, the upregulation of IGF2 can be a

criterion for selecting patients who will benefit from anti-cancer therapy targeted to the genes in Group 1 and their associated pathway components.

Group 2

Genes that exhibit differential expression in colon cancer in a pattern that correlates Interferon Induced Transmembrane (IFITM) protein family

[00218] Using the Spearman rank correlation test, the differential expression data of the IFITM family (1-8U; 1-8D; 9-27) correlated with that of 2 other genes: ITAK and BIRC3/H-IAP1 (see Table 3). The differential expression data for these genes is presented in graphical form in Figs. 17-20. This group was identified as Group 2. 1-8U/IFITM3 was previously reported as a gene differentially upregulated in ulcerative-colitis-associated colon cancer (Hisamatsu et al (1999) Cancer Research 59, 5927-5931). Genes whose expression patterns correlate with 1-8U/IFITM3 and its family members may provide a mechanism for the involvement of inflammation in colon cancer. There are at least 3 members of the IFITM family: 9-27/IFITM1, 1-8D/IFITM2 and 1-8U/IFITM3. The polynucleotides used for the detection of 1-8U/IFITM3 are within a domain that is highly conserved among the 3 members. Therefore, the upregulation detected by the corresponding microarray spots may indicate the upregulation of one or multiple members within the family. Among the genes in Group 2 are ITAK (IL-1, TNF alpha activated kinase) and BIRC3/H-IAP1 (human inhibitor of apoptosis 1). The protein products of these genes and their associated signaling pathways can be targets for small molecule drug development for anti-cancer therapy. Furthermore, the upregulation of the IFITM can be a criterion for selecting patients who will benefit from anti-cancer therapy targeted to the genes in Group 2 and their associated pathway components.

EXAMPLE 4: ANTISENSE REGULATION OF GENE EXPRESSION

[00219] The expression of the differentially expressed genes represented by the polynucleotides in the cancerous cells can be analyzed using antisense knockout technology to confirm the role and function of the gene product in tumorigenesis, *e.g.*, in promoting a metastatic phenotype.

**[00220]** A number of different oligonucleotides complementary to the mRNA generated by the differentially expressed genes identified herein can be designed as potential antisense oligonucleotides, and tested for their ability to suppress expression of the genes. Sets of antisense oligomers specific to each candidate target are designed using the sequences of the polynucleotides corresponding to a differentially expressed gene and the software program HYBsimulator Version 4 (available for Windows 95/Windows NT or for Power Macintosh, RNature, Inc. 1003 Health Sciences Road, West, Irvine, CA 92612 USA). Factors that are considered when designing antisense oligonucleotides include: 1) the secondary structure of oligonucleotides; 2) the secondary structure of the target gene; 3) the specificity with no or minimum cross-hybridization to other expressed genes; 4) stability; 5) length and 6) terminal GC content. The antisense oligonucleotide is designed so that it will hybridize to its target sequence under conditions of high stringency at physiological temperatures (*e.g.*, an optimal temperature for the cells in culture to provide for hybridization in the cell, *e.g.*, about 37°C), but with minimal formation of homodimers.

**[00221]** Using the sets of oligomers and the HYBsimulator program, three to ten antisense oligonucleotides and their reverse controls are designed and synthesized for each candidate mRNA transcript, which transcript is obtained from the gene corresponding to the target polynucleotide sequence of interest. Once synthesized and quantitated, the oligomers are screened for efficiency of a transcript knock-out in a panel of cancer cell lines. The efficiency of the knock-out is determined by analyzing mRNA levels using lightcycler quantification. The oligomers that resulted in the highest level of transcript knock-out, wherein the level was at least about 50%, preferably about 80-90%, up to 95% or more up to undetectable message, are selected for use in a cell-based proliferation assay, an anchorage independent growth assay, and an apoptosis assay.

**[00222]** The ability of each designed antisense oligonucleotide to inhibit gene expression is tested through transfection into SW620 colon carcinoma cells. For each transfection mixture, a carrier molecule (such as a lipid, lipid derivative, lipid-like molecule, cholesterol, cholesterol derivative, or cholesterol-like molecule) is prepared to a working

concentration of 0.5 mM in water, sonicated to yield a uniform solution, and filtered through a 0.45  $\mu$ m PVDF membrane. The antisense or control oligonucleotide is then prepared to a working concentration of 100  $\mu$ M in sterile Millipore water. The oligonucleotide is further diluted in OptiMEM™ (Gibco/BRL), in a microfuge tube, to 2  $\mu$ M, or approximately 20  $\mu$ g oligo/ml of OptiMEM™. In a separate microfuge tube, the carrier molecule, typically in the amount of about 1.5-2 nmol carrier/ $\mu$ g antisense oligonucleotide, is diluted into the same volume of OptiMEM™ used to dilute the oligonucleotide. The diluted antisense oligonucleotide is immediately added to the diluted carrier and mixed by pipetting up and down. Oligonucleotide is added to the cells to a final concentration of 30 nM.

[00223] The level of target mRNA that corresponds to a target gene of interest in the transfected cells is quantitated in the cancer cell lines using the Roche LightCycler™ real-time PCR machine. Values for the target mRNA are normalized versus an internal control (e.g., beta-actin). For each 20  $\mu$ l reaction, extracted RNA (generally 0.2-1  $\mu$ g total) is placed into a sterile 0.5 or 1.5 ml microcentrifuge tube, and water is added to a total volume of 12.5  $\mu$ l. To each tube is added 7.5  $\mu$ l of a buffer/enzyme mixture, prepared by mixing (in the order listed) 2.5  $\mu$ l H<sub>2</sub>O, 2.0  $\mu$ l 10X reaction buffer, 10  $\mu$ l oligo dT (20 pmol), 1.0  $\mu$ l dNTP mix (10 mM each), 0.5  $\mu$ l RNasin® (20u) (Ambion, Inc., Hialeah, FL), and 0.5  $\mu$ l MMLV reverse transcriptase (50u) (Ambion, Inc.). The contents are mixed by pipetting up and down, and the reaction mixture is incubated at 42°C for 1 hour. The contents of each tube are centrifuged prior to amplification.

[00224] An amplification mixture is prepared by mixing in the following order: 1X PCR buffer II, 3 mM MgCl<sub>2</sub>, 140  $\mu$ M each dNTP, 0.175 pmol each oligo, 1:50,000 dil of SYBR® Green, 0.25 mg/ml BSA, 1 unit *Taq* polymerase, and H<sub>2</sub>O to 20  $\mu$ l. (PCR buffer II is available in 10X concentration from Perkin-Elmer, Norwalk, CT). In 1X concentration it contains 10 mM Tris pH 8.3 and 50 mM KCl. SYBR® Green (Molecular Probes, Eugene, OR) is a dye which fluoresces when bound to double stranded DNA. As double stranded PCR product is produced during amplification, the fluorescence from

SYBR® Green increases. To each 20 µl aliquot of amplification mixture, 2 µl of template RT is added, and amplification is carried out according to standard protocols. The results are expressed as the percent decrease in expression of the corresponding gene product relative to non-transfected cells, vehicle-only transfected (mock-transfected) cells, or cells transfected with reverse control oligonucleotides.

EXAMPLE 5: EFFECT OF EXPRESSION ON PROLIFERATION

- [00225] The effect of gene expression on the inhibition of cell proliferation can be assessed in metastatic breast cancer cell lines (MDA-MB-231 ("231")); SW620 colon colorectal carcinoma cells; SKOV3 cells (a human ovarian carcinoma cell line); or LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 prostate cancer cells.
- [00226] Cells are plated to approximately 60-80% confluency in 96-well dishes. Antisense or reverse control oligonucleotide is diluted to 2 µM in OptiMEM™. The oligonucleotide-OptiMEM™ can then be added to a delivery vehicle, which delivery vehicle can be selected so as to be optimized for the particular cell type to be used in the assay. The oligo/delivery vehicle mixture is then further diluted into medium with serum on the cells. The final concentration of oligonucleotide for all experiments can be about 300 nM.
- [00227] Antisense oligonucleotides are prepared as described above (see Example 4). Cells are transfected overnight at 37°C and the transfection mixture is replaced with fresh medium the next morning. Transfection is carried out as described above in Example 4.
- [00228] Those antisense oligonucleotides that result in inhibition of proliferation of SW620 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous colon cells. Those antisense oligonucleotides that inhibit proliferation in SKOV3 cells represent genes that play a role in production or maintenance of the cancerous phenotype in cancerous breast cells. Those antisense oligonucleotides that result in inhibition of proliferation of MDA-MB-231 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous ovarian cells. Those antisense oligonucleotides that inhibit proliferation in LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 cells represent genes that



play a role in production or maintenance of the cancerous phenotype in cancerous prostate cells.

EXAMPLE 6: EFFECT OF GENE EXPRESSION ON CELL MIGRATION

- [00229] The effect of gene expression on the inhibition of cell migration can be assessed in SW620 colon cancer cells using static endothelial cell binding assays, non-static endothelial cell binding assays, and transmigration assays.
- [00230] For the static endothelial cell binding assay, antisense oligonucleotides are prepared as described above (see Example 4). Two days prior to use, colon cancer cells (CaP) are plated and transfected with antisense oligonucleotide as described above (see Examples 4 and 5). On the day before use, the medium is replaced with fresh medium, and on the day of use, the medium is replaced with fresh medium containing 2  $\mu$ M CellTracker green CMFDA (Molecular Probes, Inc.) and cells are incubated for 30 min. Following incubation, CaP medium is replaced with fresh medium (no CMFDA) and cells are incubated for an additional 30-60 min. CaP cells are detached using CMF PBS/2.5 mM EDTA or trypsin, spun and resuspended in DMEM/1% BSA/ 10 mM HEPES pH 7.0. Finally, CaP cells are counted and resuspended at a concentration of  $1 \times 10^6$  cells/ml.
- [00231] Endothelial cells (EC) are plated onto 96-well plates at 40-50% confluence 3 days prior to use. On the day of use, EC are washed 1X with PBS and 50 $\lambda$  DMDM/1%BSA/10mM HEPES pH 7 is added to each well. To each well is then added 50K (50 ) CaP cells in DMEM/1% BSA/ 10mM HEPES pH 7. The plates are incubated for an additional 30 min and washed 5X with PBS containing  $Ca^{++}$  and  $Mg^{++}$ . After the final wash, 100  $\mu$ L PBS is added to each well and fluorescence is read on a fluorescent plate reader (Ab492/Em 516 nm).
- [00232] For the non-static endothelial cell binding assay, CaP are prepared as described above. EC are plated onto 24-well plates at 30-40% confluence 3 days prior to use. On the day of use, a subset of EC are treated with cytokine for 6 hours then washed 2X with PBS. To each well is then added 150-200K CaP cells in DMEM/1% BSA/ 10mM HEPES pH 7. Plates are placed on a rotating shaker (70 RPM) for 30 min and then washed 3X

with PBS containing  $\text{Ca}^{++}$  and  $\text{Mg}^{++}$ . After the final wash, 500  $\mu\text{L}$  PBS is added to each well and fluorescence is read on a fluorescent plate reader (Ab492/Em 516 nm).

[00233] For the transmigration assay, CaP are prepared as described above with the following changes. On the day of use, CaP medium is replaced with fresh medium containing 5  $\mu\text{M}$  CellTracker green CMFDA (Molecular Probes, Inc.) and cells are incubated for 30 min. Following incubation, CaP medium is replaced with fresh medium (no CMFDA) and cells are incubated for an additional 30-60 min. CaP cells are detached using CMF PBS/2.5 mM EDTA or trypsin, spun and resuspended in EGM-2-MV medium. Finally, CaP cells are counted and resuspended at a concentration of  $1 \times 10^6$  cells/ml.

[00234] EC are plated onto FluorBlok transwells (BD Biosciences) at 30-40% confluence 5-7 days before use. Medium is replaced with fresh medium 3 days before use and on the day of use. To each transwell is then added 50K labeled CaP. 30 min prior to the first fluorescence reading, 10  $\mu\text{g}$  of FITC-dextran (10K MW) is added to the EC plated filter. Fluorescence is then read at multiple time points on a fluorescent plate reader (Ab492/Em 516 nm).

[00235] Those antisense oligonucleotides that result in inhibition of binding of SW620 colon cancer cells to endothelial cells indicate that the corresponding gene plays a role in the production or maintenance of the cancerous phenotype in cancerous colon cells. Those antisense oligonucleotides that result in inhibition of endothelial cell transmigration by SW620 colon cancer cells indicate that the corresponding gene plays a role in the production or maintenance of the cancerous phenotype in cancerous colon cells.

#### EXAMPLE 7: EFFECT OF GENE EXPRESSION ON COLONY FORMATION

[00236] The effect of gene expression upon colony formation of SW620 cells, SKOV3 cells, MD-MBA-231 cells, LNCaP cells, PC3 cells, 22Rv1 cells, MDA-PCA-2b cells, and DU145 cells can be tested in a soft agar assay. Soft agar assays are conducted by first establishing a bottom layer of 2 ml of 0.6% agar in media plated fresh within a few hours of layering on the cells. The cell layer is formed on the bottom layer by removing cells transfected as described above from plates using 0.05% trypsin and washing twice in

media. The cells are counted in a Coulter counter, and resuspended to  $10^6$  per ml in media. 10  $\mu$ l aliquots are placed with media in 96-well plates (to check counting with WST1), or diluted further for the soft agar assay. 2000 cells are plated in 800  $\mu$ l 0.4% agar in duplicate wells above 0.6% agar bottom layer. After the cell layer agar solidifies, 2 ml of media is dribbled on top and antisense or reverse control oligo (produced as described in Example 4) is added without delivery vehicles. Fresh media and oligos are added every 3-4 days. Colonies form in 10 days to 3 weeks. Fields of colonies are counted by eye. Wst-1 metabolism values can be used to compensate for small differences in starting cell number. Larger fields can be scanned for visual record of differences.

[00237] Those antisense oligonucleotides that result in inhibition of colony formation of SW620 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous colon cells. Those antisense oligonucleotides that inhibit colony formation in SKOV3 cells represent genes that play a role in production or maintenance of the cancerous phenotype in cancerous breast cells. Those antisense oligonucleotides that result in inhibition of colony formation of MDA-MB-231 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous ovarian cells. Those antisense oligonucleotides that inhibit colony formation in LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 cells represent genes that play a role in production or maintenance of the cancerous phenotype in cancerous prostate cells.

EXAMPLE 8: INDUCTION OF CELL DEATH UPON DEPLETION OF POLYPEPTIDES BY DEPLETION OF MRNA ("ANTISENSE KNOCKOUT")

[00238] In order to assess the effect of depletion of a target message upon cell death, SW620 cells, or other cells derived from a cancer of interest, can be transfected for proliferation assays. For cytotoxic effect in the presence of cisplatin (cis), the same protocol is followed but cells are left in the presence of 2  $\mu$ M drug. Each day, cytotoxicity is monitored by measuring the amount of LDH enzyme released in the medium due to membrane damage. The activity of LDH is measured using the Cytotoxicity Detection Kit

from Roche Molecular Biochemicals. The data is provided as a ratio of LDH released in the medium vs. the total LDH present in the well at the same time point and treatment (rLDH/tLDH). A positive control using antisense and reverse control oligonucleotides for BCL2 (a known anti-apoptotic gene) is included; loss of message for BCL2 leads to an increase in cell death compared with treatment with the control oligonucleotide (background cytotoxicity due to transfection).

EXAMPLE 9: FUNCTIONAL ANALYSIS OF GENE PRODUCTS DIFFERENTIALLY EXPRESSED IN COLON CANCER IN PATIENTS

[00239] The gene products of sequences of a gene differentially expressed in cancerous cells can be further analyzed to confirm the role and function of the gene product in tumorigenesis, *e.g.*, in promoting or inhibiting development of a metastatic phenotype. For example, the function of gene products corresponding to genes identified herein can be assessed by blocking function of the gene products in the cell. For example, where the gene product is secreted or associated with a cell surface membrane, blocking antibodies can be generated and added to cells to examine the effect upon the cell phenotype in the context of, for example, the transformation of the cell to a cancerous, particularly a metastatic, phenotype. In order to generate antibodies, a clone corresponding to a selected gene product is selected, and a sequence that represents a partial or complete coding sequence is obtained. The resulting clone is expressed, the polypeptide produced isolated, and antibodies generated. The antibodies are then combined with cells and the effect upon tumorigenesis assessed.

[00240] Where the gene product of the differentially expressed genes identified herein exhibits sequence homology to a protein of known function (*e.g.*, to a specific kinase or protease) and/or to a protein family of known function (*e.g.*, contains a domain or other consensus sequence present in a protease family or in a kinase family), then the role of the gene product in tumorigenesis, as well as the activity of the gene product, can be examined using small molecules that inhibit or enhance function of the corresponding protein or protein family.

[00241] Additional functional assays include, but are not necessarily limited to, those that analyze the effect of expression of the corresponding gene upon cell cycle and cell migration. Methods for performing such assays are well known in the art.

EXAMPLE 10: CONTIG ASSEMBLY AND ADDITIONAL GENE CHARACTERIZATION

[00242] The sequences of the polynucleotides provided in the present invention can be used to extend the sequence information of the gene to which the polynucleotides correspond (*e.g.*, a gene, or mRNA encoded by the gene, having a sequence of the polynucleotide described herein). This expanded sequence information can in turn be used to further characterize the corresponding gene, which in turn provides additional information about the nature of the gene product (*e.g.*, the normal function of the gene product). The additional information can serve to provide additional evidence of the gene product's use as a therapeutic target, and provide further guidance as to the types of agents that can modulate its activity.

[00243] In one example, a contig is assembled using a sequence of a polynucleotide of the present invention, which is present in a clone. A "contig" is a contiguous sequence of nucleotides that is assembled from nucleic acid sequences having overlapping (*e.g.*, shared or substantially similar) sequence information. The sequences of publicly-available ESTs (Expressed Sequence Tags) and the sequences of various clones from several cDNA libraries synthesized at Chiron can be used in the contig assembly.

[00244] The contig is assembled using the software program Sequencher, version 4.05, according to the manufacturer's instructions and an overview alignment of the contiged sequences is produced. The sequence information obtained in the contig assembly can then be used to obtain a consensus sequence derived from the contig using the Sequencher program. The consensus sequence is used as a query sequence in a TeraBLASTN search of the DGTI DoubleTwist Gene Index (DoubleTwist, Inc., Oakland, CA), which contains all the EST and non-redundant sequence in public databases.

[00245] Through contig assembly and the use of homology searching software programs, the sequence information provided herein can be readily extended to confirm, or confirm a

predicted, gene having the sequence of the polynucleotides described in the present invention. Further the information obtained can be used to identify the function of the gene product of the gene corresponding to the polynucleotides described herein. While not necessary to the practice of the invention, identification of the function of the corresponding gene, can provide guidance in the design of therapeutics that target the gene to modulate its activity and modulate the cancerous phenotype (*e.g.*, inhibit metastasis, proliferation, and the like).

**[00246]** While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the invention.

Table 1

Pt ID	Path ID	Grp	Anatom Loc	Size	Grade	Histo Grade	Local Invasion	Lymph Met	Lymph Met Incid	Reg Lymph Grade	Dist Met & Loc	Dist Met Grade	Comment
15	21	III	Ascending colon	4.0	T3	G2	Extending into subserosal adipose tissue	Pos	3/8	N1	Neg	MX	invasive adenocarcinoma, moderately differentiated; focal perineural invasion is seen
52	71	II	Cecum	9.0	T3	G3	Invasion through muscularis propria, subserosal involvement; ileocecal valve involvement	Neg	0/12	N0	Neg	M0	Hyperplastic polyp in appendix.
121	140	II	Sigmoid	6	T4	G2	Invasion of muscularis propria into serosa, involving submucosa of urinary bladder	Neg	0/34	N0	Neg	M0	Perineural invasion; donut anastomosis Neg. One tubulovillous and one tubular adenoma with no high grade dysplasia.

Table 1

Pt ID	Path ID	Grp	Anatom Loc	Size	Grade	Histo Grade	Local Invasion	Lymph Met	Lymph Met Incid	Reg Lymph Grade	Dist Met & Loc	Dist Met Grade	Comment
125	144	II	Cecum	6	T3	G2	Invasion through the muscularis propria into subserosal adipose tissue. Ileocecal junction.	Neg	0/19	N0	Neg	M0	patient history of metastatic melanoma
128	147	III	Transverse colon	5.0	T3	G2	Invasion of muscularis propria into pericolonic fat	Pos	1/5	N1	Neg	M0	
130	149		Splenic flexure	5.5	T3		through wall and into surrounding adipose tissue	Pos	10/24	N2	Neg	M1	
133	152	II	Rectum	5.0	T3	G2	Invasion through muscularis propria into non-peritonealized pericolic tissue; gross configuration is annular.	Neg	0/9	N0	Neg	M0	Small separate tubular adenoma (0.4 cm)
141	160	IV	Cecum	5.5	T3	G2	Invasion of muscularis propria into pericolonic adipose tissue, but not through serosa. Arising from tubular adenoma.	Pos	7/21	N2	Pos - Liver	M1	Perineural invasion identified adjacent to metastatic adenocarcinoma.



Table 1

Pt ID	Path ID	Grp	Anatom Loc	Size	Grade	Histo Grade	Local Invasion	Lymph Met	Lymph Met Incid	Reg Lymph Grade	Dist Met & Loc	Dist Met Grade	Comment
156	175	III	Hepatic flexure	3.8	T3	G2	Invasion through muscularis propria into subserosa/pericolonic adipose, no serosal involvement. Gross configuration annular.	Pos	2/13	N1	Neg	M0	Separate tubulovillous and tubular adenomas
228	247	III	Rectum	5.8	T3	G2 to G3	Invasion through muscularis propria to involve subserosal, perirectal adipose, and serosa	Pos	1/8	N1	Neg	MX	Hyperplastic polyps
264	283	II	Ascending colon	5.5	T3	G2	Invasion through muscularis propria into subserosal adipose tissue.	Neg	0/10	N0	Neg	M0	Tubulovillous adenoma with high grade dysplasia
266	285	III	Transverse colon	9	T3	G2	Invades through muscularis propria to involve pericolonic adipose, extends to serosa.	Neg	0/15	N1	Pos - Mesenteric deposit	MX	
268	287	I	Cecum	6.5	T2	G2	Invades full thickness of muscularis propria, but mesenteric adipose free of malignancy	Neg	0/12	N0	Neg	M0	

Table 1

Pt ID	Path ID	Grp	Anatom Loc	Size	Grade	Histo Grade	Local Invasion	Lymph Met	Lymph Met Incid	Reg Lymph Grade	Dist Met & Loc	Dist Met Grade	Comment
278	297	III	Rectum	4	T3	G2	Invasion into perirectal adipose tissue.	Pos	7/10	N2	Neg	M0	Descending colon polyps, no HGD or carcinoma identified..
295	314	II	Ascending colon	5.0	T3	G2	Invasion through muscularis propria into percolic adipose tissue.	Neg	0/12	N0	Neg	M0	Melanosis coli and diverticular disease.
296	315	III	Cecum	5.5	T3	G2	Invasion through muscularis propria and invades pericolic adipose tissue. Ileocecal junction.	Pos	2/12	N1	Neg	M0	Tubulovillous adenoma (2.0 cm) with no high grade dysplasia. Neg. liver biopsy.
339	358	II	Rectosigmoid	6	T3	G2	Extends into perirectal fat but does not reach serosa	Neg	0/6	N0	Neg	M0	1 hyperplastic polyp identified
341	360	II	Ascending colon	2 cm invasive	T3	G2	Invasion through muscularis propria to involve pericolic fat. Arising from villous adenoma.	Neg	0/4	N0	Neg	MX	
356	375	II	Sigmoid	6.5	T3	G2	Through colon wall into subserosal adipose tissue. No serosal spread seen.	Neg	0/4	N0	Neg	M0	

Table 1

Pt ID	Path ID	Grp	Anatom Loc	Size	Grade	Histo Grade	Local Invasion	Lymph Met	Lymph Met Incid	Reg Lymph Grade	Dist Met & Loc	Dist Met Grade	Comment
392	444	IV	Ascending colon	2	T3	G2	Invasion through muscularis propria into subserosal adipose tissue, not serosa.	Pos	1/6	N1	Pos - Liver	M1	Tumor arising at prior ileocolic surgical anastomosis.
393	445	II	Cecum	6.0	T3	G2	Cecum, invades through muscularis propria to involve subserosal adipose tissue but not serosa.	Neg	0/21	N0	Neg	M0	
413	465	IV	Cecum	4.8	T3	G2	Invasive through muscularis to involve periserosal fat; abutting ileocecal junction.	Neg	0/7	N0	Pos - Liver	M1	redagnosis of oophorectomy path to metastatic colon cancer.
517	395	IV	Sigmoid	3	T3	G2	penetrates muscularis propria, involves pericolic fat.	Pos	6/6	N2	Neg	M0	No mention of distant met in report
546	565	IV	Ascending colon	5.5	T3	G2	Invasion through muscularis propria extensively through submucosal and extending to serosa.	Pos	6/12	N2	Pos - Liver	M1	
577	596	II	Cecum	11.5	T3	G2	Invasion through the bowel wall, into subserosal adipose. Serosal surface free of tumor.	Neg	0/58	N0	Neg	M0	Appendix dilated and fibrotic, but not involved by tumor

Table 1

Pt ID	Path ID	Grp	Anatom Loc	Size	Grade	Histo Grade	Local Invasion	Lymph Met	Lymph Met Incid	Reg Lymph Grade	Dist Met & Loc	Dist Met Grade	Comment
784	803	IV	Ascending colon	3.5	T3	G3	through muscularis propria into pericolic soft tissues	Pos	5/17	N2	Pos - Liver	M1	invasive poorly differentiated adenocarcinoma
786	805	IV	Descending colon	9.5	T3	G2	through muscularis propria into pericolic fat, but not at serosal surface	Neg	0/12	N0	Pos - Liver	M1	moderately differentiated invasive adenocarcinoma
791	810	IV	Ascending colon	5.8	T3	G3	Through the muscularis propria into pericolic fat	Pos	13/25	N2	Pos - Liver	M1	poorly differentiated invasive colonic adenocarcinoma
888	908	IV	Ascending colon	2.0	T2	G1	Into muscularis propria	Pos	3/21	N0	Pos - Liver	M1	well to moderately differentiated adenocarcinoma; this patient has tumors of the ascending colon and the sigmoid colon
889	909	IV	Cecum	4.8	T3	G2	Through muscularis propria into subserosal tissue	Pos	1/4	N1	Pos - Liver	M1	moderately differentiated adenocarcinoma

Table 1

Pt ID	Path ID	Grp	Anatom Loc	Size	Grade	Histo Grade	Local Invasion	Lymph Met	Lymph Met Incid	Reg Lymph Grade	Dist Met & Loc	Dist Met Grade	Comment
890	910	IV	Ascending colon		T3	G2	Through muscularis propria into subserosa.	Pos	11/15	N2	Pos - Liver	M1	

Table 2

SEQ ID NO	Spot ID	Clone ID	Grp	Gene	GBHit	GBDesc	GBScore
1	33669	RG:26148:Order7TM01:C06	1	IGF2	X07868	Human DNA for insulin-like growth factor II (IGF-2); exon 7 and additional ORF	2.1E-35
2	32956	RG:240381:Order7TM20:G11	1	IGF2	X03427	Homo sapiens IGF-II gene, exon 5	7.4E-186
3	17167	RG:730402:10010:H01	1	TTK	BC000633	Homo sapiens, TTK protein kinase, clone MGC:865 IMAGE:3343925, mRNA, complete cds	2.1E-38
4	21711	RG:1674098:10014:H01	1	MARCK S	D10522	Homo sapiens mRNA for 80K-L protein, complete cds	4E-148
5	29171	035JN025.C12	1	FLJ2206 6	AK025719	Homo sapiens cDNA: FLJ22066 fis, clone HEP10611	0
6	30566	RG:432087:Order7TM26:D02	1	FLJ2206 6	AK025719	Homo sapiens cDNA: FLJ22066 fis, clone HEP10611	0
7	10638	I:1644648:07B01:G04	1	NQO2	U07736	Human quinone oxidoreductase2 (NQO2) gene, exon 7, complete cds	1.6E-171
8	8491	I:2594080:05A01:F01	1	FHL3	BC001351	Homo sapiens, Similar to four and a half LIM domains 3, clone MGC:8696 IMAGE:2964682, mRNA, compl	2.6E-34
9	27092	035Jn031.C09	1	MGC:29 604	BC019103	Homo sapiens, clone MGC:29604 IMAGE:5021401, mRNA, complete cds	1E-300
10	10394	I:1450639:03B02:E09	1	CETN2	BC005334	Homo sapiens, centrin, EF-hand protein, 2, clone MGC:12421 IMAGE:3961448, mRNA, complete cds	1.1E-190
11	3295	M00008083D:D06	1	CGI-148 protein	AF223467	Homo sapiens NPD008 protein (NPD008) mRNA, complete cds	2.5E-157
12	30831	RG:301734:Order7TM22:H02	1	KIP2	AB012955	Homo sapiens mRNA for KIP2, complete cds	5.8E-252
13	19871	RG:196236:10006:H11	1	FGFR4	AF359246	Homo sapiens fibroblast growth factor receptor 4 variant mRNA, complete cds	5E-249
14	30858	RG:359021:Order7TM24:F02	1	BBS2	AF342736	Homo sapiens BBS2 (BBS2) mRNA, complete cds	1E-100
15	17168	RG:1320327:10012:H01	1	OGG1	Y11731	H.sapiens mRNA for DNA glycosylase	1E-300

Table 2

SEQ ID NO	Spot ID	Clone ID	Grp	Gene	GBHit	GBDesc	GBScore
16	17487	RG:341475:10008:H01	1	MAPKA PK2	NM_03296 0	Homo sapiens mitogen-activated protein kinase-activated protein kinase 2 (MAPKAPK2), transcript variant	1E-300
17	18942	RG:1895716:10015:G09	2	ITAK	AC007055	AC007055 Homo sapiens chromosome 14 clone BAC 201F1 map 14q24.3, complete sequence	3.00E-94
18	17365	I:504786:14A02:C07	2	1-8U; 1-8D; 9-27	BC006794	Homo sapiens, Similar to interferon induced transmembrane protein 3 (1-8U), clone MGC:5225 IMAGE:	6.4E-295
19	21144	M00055353D:A04	2	1-8U; 1-8D; 9-27	BC006794	Homo sapiens, Similar to interferon induced transmembrane protein 3 (1-8U), clone MGC:5225 IMAGE:	1.1E-156
20	11573	I:1513214:04A01:C11	2	BIRC3	U45878	Human inhibitor of apoptosis protein 1 mRNA, complete cds	2.5E-157

Table 3

Spot ID	Grp	Gene	RATIO015	RATIO052	RATIO121	RATIO125
3295	1	CGI-148 protein	0.603	0.569	1.420	1.000
8491	1	FHL3	1.000	1.000	10.786	6.347
10394	1	CETN2	1.000	1.000	3.335	1.000
10638	1	NQO2	1.000	1.000	2.522	1.720
17167	1	TTK	1.000	1.000	5.053	1.000
17168	1	OGG1	1.389	1.000	1.736	1.000
17487	1	MAPKAPK2	1.000	1.000	39.041	1.000
19871	1	FGFR4	1.000	1.000	4.040	0.760
21711	1	MARCKS	1.000	1.000	21.440	1.294
27092	1	MGC:29604	1.806	2.418	5.831	2.114
29171	1	FLJ22066	1.000	1.000	184.016	0.728
30566	1	FLJ22066	1.000	1.000	163.068	1.000
30831	1	KIP2	0.723	1.000	2.349	1.000
30858	1	BBS2	1.304	0.745	1.907	1.678
32956	1	IGF2	1.105	1.000	20.747	1.000
33669	1	IGF2	0.592	0.381	21.028	1.195
11573	2	BIRC3	1.698	2.791	0.825	1.319
17365	2	1-8U; 1-8D; 9-27	3.113	2.893	1.229	4.848
18942	2	ITAK	4.489	7.386	1.000	6.655
21144	2	1-8U; 1-8D; 9-27	5.520	22.946	1.000	5.929



Table 3

Spot ID	Grp	Gene	RATIO128	RATIO130	RATIO133	RATIO141
3295	1	CGI-148 protein	1.347	0.544	1.000	0.663
8491	1	FHL3	4.580	2.918	5.331	1.000
10394	1	CETN2	2.493	2.450	1.000	1.000
10638	1	NQO2	2.495	1.000	1.748	1.000
17167	1	TTK	5.484	1.000	1.000	1.000
17168	1	OGG1	2.525	1.000	2.339	1.000
17487	1	MAPKAPK2	26.551	1.000	54.030	0.657
19871	1	FGFR4	3.246	1.000	4.017	1.859
21711	1	MARCKS	10.369	1.000	20.040	1.000
27092	1	MGC:29604	11.273	1.821	9.841	1.413
29171	1	FLJ22066	52.758	0.849	145.030	1.000
30566	1	FLJ22066	53.616	1.000	1.000	1.000
30831	1	KIP2	1.972	1.000	1.000	1.437
30858	1	BBS2	2.686	0.525	1.877	1.000
32956	1	IGF2	10.458	1.000	1.000	1.000
33669	1	IGF2	16.876	0.334	25.468	0.720
11573	2	BIRC3	1.264	1.587	1.986	0.408
17365	2	1-8U; 1-8D; 9-27	3.307	4.004	9.166	1.000
18942	2	ITAK	4.507	5.485	12.390	1.000
21144	2	1-8U; 1-8D; 9-27	3.918	7.337	8.908	1.182

Table 3

Spot ID	Grp	Gene	RATIO156	RATIO228	RATIO264	RATIO266
3295	1	CGI-148 protein	0.400	0.579	0.599	0.302
8491	1	FHL3	2.771	1.000	1.000	1.000
10394	1	CETN2	2.130	1.000	1.000	1.000
10638	1	NQO2	2.018	1.000	1.000	1.000
17167	1	TTK	1.000	1.000	1.724	1.515
17168	1	OGG1	1.162	1.000	1.584	1.332
17487	1	MAPKAPK2	0.116	1.000	1.000	1.206
19871	1	FGFR4	0.224	1.619	1.992	1.000
21711	1	MARCKS	1.000	1.000	1.000	1.192
27092	1	MGC:29604	2.385	1.927	3.330	2.678
29171	1	FLJ22066	0.015	1.000	1.760	1.000
30566	1	FLJ22066	0.083	1.596	1.430	1.000
30831	1	KIP2	0.626	0.672	0.952	1.000
30858	1	BBS2	0.251	1.393	1.547	1.431
32956	1	IGF2	0.476	1.000	1.000	1.000
33669	1	IGF2	0.049	0.566	0.380	0.196
11573	2	BIRC3	1.504	1.000	1.645	1.000
17365	2	1-8U; 1-8D; 9-27	1.769	2.633	7.263	7.775
18942	2	ITAK	2.281	4.106	10.286	11.733
21144	2	1-8U; 1-8D; 9-27	1.706	5.027	8.086	8.148

Table 3

Spot ID	Grp	Gene	RATIO268	RATIO278	RATIO295	RATIO296
3295	1	CGI-148 protein	1.000	1.270	1.000	0.484
8491	1	FHL3	12.583	4.691	1000.000	1000.000
10394	1	CETN2	3.463	1.000	1.000	1000.000
10638	1	NQO2	3.325	1.697	1.000	1000.000
17167	1	TTK	1.000	1.000	1.000	1000.000
17168	1	OGG1	2.564	2.024	1.600	1.551
17487	1	MAPKAPK2	43.580	23.642	2.085	1.000
19871	1	FGFR4	4.407	3.989	1000.000	1.000
21711	1	MARCKS	13.283	1.000	2.161	1.000
27092	1	MGC:29604	10.984	9.190	4.226	8.035
29171	1	FLJ22066	186.617	83.660	4.242	1000.000
30566	1	FLJ22066	108.781	51.686	1.000	1.000
30831	1	KIP2	1.000	2.848	1.000	1.000
30858	1	BBS2	2.272	1.440	1.000	1.000
32956	1	IGF2	32.991	3.788	1.000	1.000
33669	1	IGF2	14.331	4.654	0.298	0.237
11573	2	BIRC3	1.283	1.667	1.408	2.084
17365	2	1-8U; 1-8D; 9-27	4.152	4.770	3.064	2.220
18942	2	ITAK	6.840	1.000	11.385	1.000
21144	2	1-8U; 1-8D; 9-27	3.902	7.228	5.159	1.000

Table 3

Spot ID	Grp	Gene	RATIO339	RATIO341	RATIO356	RATIO392
3295	1	CGI-148 protein	0.561	1.000	0.503	0.816
8491	1	FHL3	3.136	7.320	1.000	1.000
10394	1	CETN2	1.000	4.065	1000.000	1.000
10638	1	NQO2	1.000	3.036	1.000	1.000
17167	1	TTK	1.000	5.355	1.000	1.000
17168	1	OGG1	0.739	1.999	1.000	2.116
17487	1	MAPKAPK2	0.545	18.309	1.556	51.316
19871	1	FGFR4	1.324	2.494	1.000	2.284
21711	1	MARCKS	0.638	1.000	1.000	32.171
27092	1	MGC:29604	0.757	14.757	7.284	12.948
29171	1	FLJ22066	0.303	102.601	1.000	218.198
30566	1	FLJ22066	0.530	50.061	1.000	264.417
30831	1	KIP2	1.000	2.521	1.000	1.997
30858	1	BBS2	1.000	2.180	0.519	3.152
32956	1	IGF2	1.565	10.202	1.475	25.053
33669	1	IGF2	0.508	11.442	0.412	24.283
11573	2	BIRC3	1.000	1.000	1.000	1.199
17365	2	1-8U; 1-8D; 9-27	1.374	1.808	3.636	9.985
18942	2	ITAK	1.892	1.690	12.611	16.163
21144	2	1-8U; 1-8D; 9-27	2.787	1.569	10.080	18.239

Table 3

Spot ID	Grp	Gene	RATIO393	RATIO413	RATIO517	RATIO546
3295	1	CGI-148 protein	0.692	0.649	0.200	1.000
8491	1	FHL3	13.185	1.000	1000.000	3.131
10394	1	CETN2	3.015	1.000	1.000	1.000
10638	1	NQO2	2.850	1.000	1.000	1.000
17167	1	TTK	5.355	1.000	1.000	1.000
17168	1	OGG1	1.694	1.000	1.000	1.000
17487	1	MAPKAPK2	43.253	0.516	1.412	0.813
19871	1	FGFR4	4.041	1.000	3.005	2.185
21711	1	MARCKS	26.574	0.814	1.000	1.000
27092	1	MGC:29604	8.685	1.742	1.451	2.296
29171	1	FLJ22066	197.610	0.330	1.657	0.749
30566	1	FLJ22066	157.238	0.293	1.300	1.000
30831	1	KIP2	1.964	1.000	1.379	1.119
30858	1	BBS2	2.475	3.013	0.449	1.000
32956	1	IGF2	23.953	1.000	1.529	1.430
33669	1	IGF2	30.632	0.564	0.214	0.853
11573	2	BIRC3	1.768	1.000	1.485	1.000
17365	2	1-8U; 1-8D; 9-27	7.293	2.980	4.484	3.107
18942	2	ITAK	7.279	3.603	6.904	4.196
21144	2	1-8U; 1-8D; 9-27	8.395	2.839	6.176	3.328

Table 3

Spot ID	Grp	Gene	RATIO577	RATIO784	RATIO786	RATIO791
3295	1	CGI-148 protein	1.000	0.662	0.532	0.495
8491	1	FHL3	5.278	1.000	1.000	1.000
10394	1	CETN2	1.000	1.000	1.000	1.000
10638	1	NQO2	1.000	1.000	1.000	1.000
17167	1	TTK	3.158	1.092	1.898	1.000
17168	1	OGG1	1.672	1.701	1.000	1.000
17487	1	MAPKAPK2	1.000	1.000	1.000	1.000
19871	1	FGFR4	1.000	1.000	3.307	1.000
21711	1	MARCKS	1.347	1.000	1.000	1.000
27092	1	MGC:29604	3.357	1.329	2.919	3.771
29171	1	FLJ22066	1.000	1.000	1.790	1.000
30566	1	FLJ22066	1.220	2.785	1.000	1.000
30831	1	KIP2	0.753	1.972	1.000	1.000
30858	1	BBS2	0.662	1.339	1.000	0.749
32956	1	IGF2	1.600	1.430	1.713	1.486
33669	1	IGF2	0.381	0.551	0.506	0.474
11573	2	BIRC3	1.429	1.000	1.648	2.502
17365	2	1-8U; 1-8D; 9-27	4.362	1.645	4.670	8.576
18942	2	ITAK	7.792	1.000	8.475	10.189
21144	2	1-8U; 1-8D; 9-27	5.636	2.142	7.000	14.444

Table 3

Spot ID	Grp	Gene	RATIO888	RATIO889	RATIO890
3295	1	CGI-148 protein	0.574	0.483	0.711
8491	1	FHL3	1.000	1.000	5.465
10394	1	CETN2	2.970	1.000	2.848
10638	1	NQO2	1.511	1.000	2.158
17167	1	TTK	1.000	1.000	2.290
17168	1	OGG1	1.000	1.000	1.519
17487	1	MAPKAPK2	1.449	1.000	1.516
19871	1	FGFR4	1.988	0.646	4.007
21711	1	MARCKS	1.397	1.000	1.000
27092	1	MGC:29604	1.890	2.788	1.799
29171	1	FLJ22066	1.000	7.569	2.512
30566	1	FLJ22066	2.624	1.000	1.713
30831	1	KIP2	1.000	1.000	4.213
30858	1	BBS2	2.316	0.506	1.000
32956	1	IGF2	1.633	1.000	1.491
33669	1	IGF2	0.842	2.502	0.736
11573	2	BIRC3	0.781	1.314	1.000
17365	2	1-8U; 1-8D; 9-27	2.723	3.553	11.697
18942	2	ITAK	2.909	4.165	11.972
21144	2	1-8U; 1-8D; 9-27	2.712	7.659	11.467